

**ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE
DU VIVANT**

Centre de Biophysique Moléculaire

THÈSE présentée par :
D'Heygere François

soutenue le : **27 novembre 2015**

pour obtenir le grade de : **Docteur de l'université d'Orléans**

Discipline/ Spécialité : Biologie Moléculaire et Cellulaire

**Étude de la conservation de la
terminaison de la transcription Rho-
dépendante au sein de la biodiversité.**

THÈSE dirigée par :

Dr. Marc Boudvillain

Directeur de Recherche, CNRS Orléans

RAPPORTEURS :

Dr. Pascale Romby
Dr. Thierry Bizebard

Directrice de Recherche, CNRS Strasbourg
Chargé de Recherche, CNRS Paris

JURY:

Dr. Bertrand Castaing
Dr. Pascale Romby
Dr. Thierry Bizebard
Dr. Nara Figueroa-Bossi
Dr. Marc Boudvillain

Directeur de Recherche, CNRS Orléans
Directrice de Recherche, CNRS Strasbourg
Chargé de Recherche, CNRS Paris
Directrice de Recherche, CNRS Paris
Directeur de Recherche, CNRS Orléans

Remerciements

Les travaux qui ont fait l'objet de ce mémoire ont été réalisés au Centre de Biophysique Moléculaire d'Orléans sous la direction du Dr Marc Boudvillain, directeur de recherche au CNRS. Merci Marc de m'avoir accueilli dans ton équipe et de m'avoir fait découvrir le monde de l'ARN. Merci aussi d'avoir eu la patience pour combler mes lacunes, surtout au niveau de l'enzymologie et d'avoir pris le temps pour me guider. Tes conseils ont été, et seront je l'espère, de précieux atouts qui m'accompagneront durant ma carrière scientifique.

J'adresse mes sincères remerciements aux Docteurs Pascale Romby et Thierry Bizebard pour avoir accepté de juger ce travail en tant que rapporteurs. Je remercie également les Docteurs Nara Figueroa-Bossi et Bertrand Castaing de bien vouloir participer à mon jury de soutenance de thèse. Je suis vraiment honoré que tous les quatre aient consenti à me consacrer une partie de leur temps.

Je tiens à remercier tout particulièrement Annie Schwartz et Emilie Soares pour avoir su m'enseigner les bonnes pratiques à avoir dans un laboratoire ARN, car j'ai conscience que cela n'a pas toujours été facile. Merci aussi pour le soutien et les encouragements dont vous avez fait preuve.

Je remercie également les Docteurs Nara Figueroa-Bossi et Lionello Bossi pour leur collaboration ainsi que pour les conseils qu'ils ont su me prodiguer pour l'orientation de ma future carrière. Merci aussi aux membres de l'équipe de Bertrand Castaing pour leurs conseils et support technique et tout particulièrement à Franck Coste pour son aide précieuse dans la production, la purification et la caractérisation biophysique des protéines.

Je remercie chaleureusement tous mes amis, les membres de ma famille et l'ensemble des doctorants du CBM pour avoir su être là durant les moments difficiles. Un grand merci tout particulier à Tatiana Kondakova !

Abréviations

ADN : acide désoxyribonucléique

ADP : adénosine diphosphate

ADP•BeF_x : ADP-beryllium fluoride

AMP-PNP : adénosine 5'-(β,γ-imido) triphosphate

ARN : acide ribonucléique

ARNm : ARN messenger

ARNnc : ARN non-codant

ARNP : ARN polymérase

ATB : Antibiotique

ATP : adénosine triphosphate

ATPγS : adénosine 5'-O-(3-thio) triphosphate

BCM : Bicyclomycine

CET : Complexe d'élongation de la transcription

ChIP-chip : Immunoprécipitation de la chromatine sur Chip (*Chromatin Immunoprecipitation on Chip*)

CTD : Domaine carboxy-terminal (*Carboxy-terminal Domain*)

kDa : kilo Dalton

NAIM : *Nucleotide Analog Interference Mapping*

NTD : Domaine amino-terminal (*Amino-terminal Domain*)

NTP : ribonucléoside triphosphate

nt : nucléotides

pb : paire de bases

PBS : Site primaire de fixation (*Primary Binding Site*)

PNPase : Poly-nucléotide phosphorylase

Poly[rC] : Polymère de polyribocytidine

Pi : Ortho-phosphate inorganique

RBP : Protéine fixant l'ARN (*RNA Binding Protein*)

rbs : Site de fixation du ribosome (*Ribosome Binding Site*)

RMN : Résonance Magnétique Nucléaire

RNP : Ribonucléoprotéique

Rut : site de fixation pour Rho (*Rho utilization*)

SD : séquence Shine Dalgarno

SF : superfamille

sRNA : Petit ARN (*small RNA*)

SBS : Site secondaire de fixation (*Secondary Binding Site*)

trNAIP : *time-resolved Nucleotide Analog Interference Probing*

UTR : région non traduite (*Untranslated Region*)

ABREVIATIONS DES ACIDES AMINES

Acide aminé	Abréviation à 3 lettres	Abréviation à 1 lettre
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Acide aspartique	Asp	D
Cystéine	Cys	C
Acide glutamique	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Méthionine	Met	M
Phénylalanine	Phe	F
Proline	Pro	P
Sérine	Ser	S
Thréonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

ABREVIATIONS DES ACIDES NUCLEIQUES

Acide Nucléique	Type de base	Abréviation à 1 lettre
Adénosine	Purine	A
Guanine	Purine	G
Cytosine	Pyrimidine	C
Tymidine	Pyrimidine	T
Uridine	Pyrimidine	U
Adénosine / Guanine	Purines	R
Cytosine / Tymidine / Uridine	Pyrimidines	Y

Table des matières

REMERCIEMENTS.....	3
ABREVIATIONS	4
INTRODUCTION	9
CHAPITRE I : LE METABOLISME DE L'ARN	15
1. La transcription	17
<i>A. L'initiation de la transcription</i>	<i>18</i>
<i>B. L'élongation de la transcription.....</i>	<i>21</i>
<i>C. La terminaison de la transcription</i>	<i>28</i>
2. La traduction	33
<i>A. La synthèse des ARN impliqués dans la traduction.....</i>	<i>33</i>
<i>B. Ribogénèse et composition des ribosomes.....</i>	<i>37</i>
<i>C. Synthèse des protéines</i>	<i>38</i>
3. Dégradation des ARN	42
<i>A. La voie 5'-monophosphate</i>	<i>42</i>
<i>B. La voie directe</i>	<i>43</i>
<i>C. Implication des hélicases dans le dégradosome</i>	<i>44</i>
4. La régulation de l'expression génique dans la région 5'-leader des ARN _m	47
CHAPITRE II : LES ARN HELICASES	51
1. Les hélicases monomériques.....	54
<i>A. Organisation structurale des ARN hélicases "monomériques"</i>	<i>54</i>
<i>B. Mécanismes</i>	<i>56</i>
<i>C. Rôles et importance</i>	<i>59</i>
2. Les hélicases toroïdales.....	60
<i>A. Organisation structurale</i>	<i>61</i>
<i>B. Mécanismes généraux.....</i>	<i>63</i>
CHAPITRE III : LE FACTEUR RHO.....	69
1. Organisation structurale	71
<i>A. Le site primaire de fixation à l'ARN (PBS)</i>	<i>73</i>
<i>B. Le site secondaire de fixation à l'ARN (SBS)</i>	<i>74</i>
<i>C. Le site ATPase</i>	<i>78</i>

2. Les activités enzymatiques de Rho	80
A. L'activité NTPase	80
B. L'activité hélicase et importance du site Rut.....	80
3. Le mécanisme de terminaison de la transcription rho-dépendante	84
4. Les fonctions du facteur Rho	90
A. Régulation (conditionnelle) de la transcription.....	90
B. Contrôle de la stabilité du transcriptome	93
C. Protection de l'intégrité génomique	94
5. Rho et la diversité bactérienne	97
A. Spectre de la bicyclomycine et importance de Rho	97
B. La distribution du facteur Rho au sein de la diversité bactérienne.....	99
C. La conservation des mécanismes de Rho.....	101
CHAPITRE IV : RESULTATS PERSONNELS	103
1. Introduction	105
Article I: <i>Phyletic distribution and conservation of the bacterial transcription termination factor Rho</i>	109
Article II: <i>ATP-dependent motor activity of the transcription termination factor Rho from Mycobacterium tuberculosis</i>	131
Article III : <i>RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination</i>	151
2. Discussion	171
3. Conclusion/perspectives.....	176
BIBLIOGRAPHIE	179

Introduction

Mes travaux de thèse s'inscrivent dans l'étude de la régulation de la transcription bactérienne et ont été réalisés au sein de l'équipe "*Hélicases & ARN : mécanismes, ciblage et bio-mimétisme*", dirigée par le Dr Marc Boudvillain et située au Centre de Biophysique Moléculaire (UPR 4301) au CNRS à Orléans.

Les bactéries ont potentiellement colonisé toutes les niches écologiques possibles, allant des milieux intracellulaires aux cheminées thermales des grands fonds. Ainsi, le règne bactérien est constitué d'une grande variété d'espèces qui ont su s'adapter à leurs niches écologiques respectives. Ces niches peuvent correspondre à des environnements relativement stables ou, au contraire, à des environnements extrêmement variables imposant divers stress à la bactérie. Pour s'adapter rapidement à ces stress, plusieurs stratégies sont employées par ces organismes. Parmi elles, un changement dans l'expression du génome est primordial et permet d'adapter le métabolisme bactérien aux nouvelles conditions environnementales (Barria *et al.*, 2013; Michaux *et al.*, 2014; Redder *et al.*, 2015; Vakulskas *et al.*, 2015). La modulation du métabolisme de l'ARN joue un rôle particulièrement important dans ce processus d'adaptation étant donné que la plupart des ARN produits par la cellule sont impliqués dans l'expression du génome et sa régulation. Comme nous le verrons dans le Chapitre I les différentes étapes de ce métabolisme peuvent varier selon les espèces bactériennes, reflétant une adaptation au cours de leur spéciation.

Le métabolisme de l'ARN requiert des cofacteurs permettant aux molécules d'ARN d'adopter leurs conformations tridimensionnelles fonctionnelles (Anantharaman *et al.*, 2002). Parmi ces cofacteurs on retrouve de nombreuses enzymes comme des ARN chaperonnes, des endoribonucléases ou des ARN hélicases, sujets d'étude de l'équipe. Les ARN hélicases forment un groupe polyphylétique d'enzymes adoptant diverses structures tridimensionnelles et qui sont capables de provoquer le remodelage de molécules d'ARN et/ou de complexes ribonucléoprotéiques (RNP) via différents mécanismes dépendants de la présence de nucléotide-triphosphate (NTP) (Fairman-Williams *et al.*, 2010; Jankowsky & Fairman-Williams, 2010; Pyle, 2008; Singleton *et al.*, 2007). Ces enzymes sont retrouvées chez tous les organismes vivants, ainsi que chez certains virus, et participent vraisemblablement à toutes les étapes du métabolisme de l'ARN (Jankowsky & Fairman, 2007; Jankowsky & Fairman-Williams, 2010; Jarmoskaite & Russell, 2014; Rocak & Linder, 2004). Un bref exposé des principales caractéristiques des ARN hélicases est proposé dans le Chapitre II de ce manuscrit.

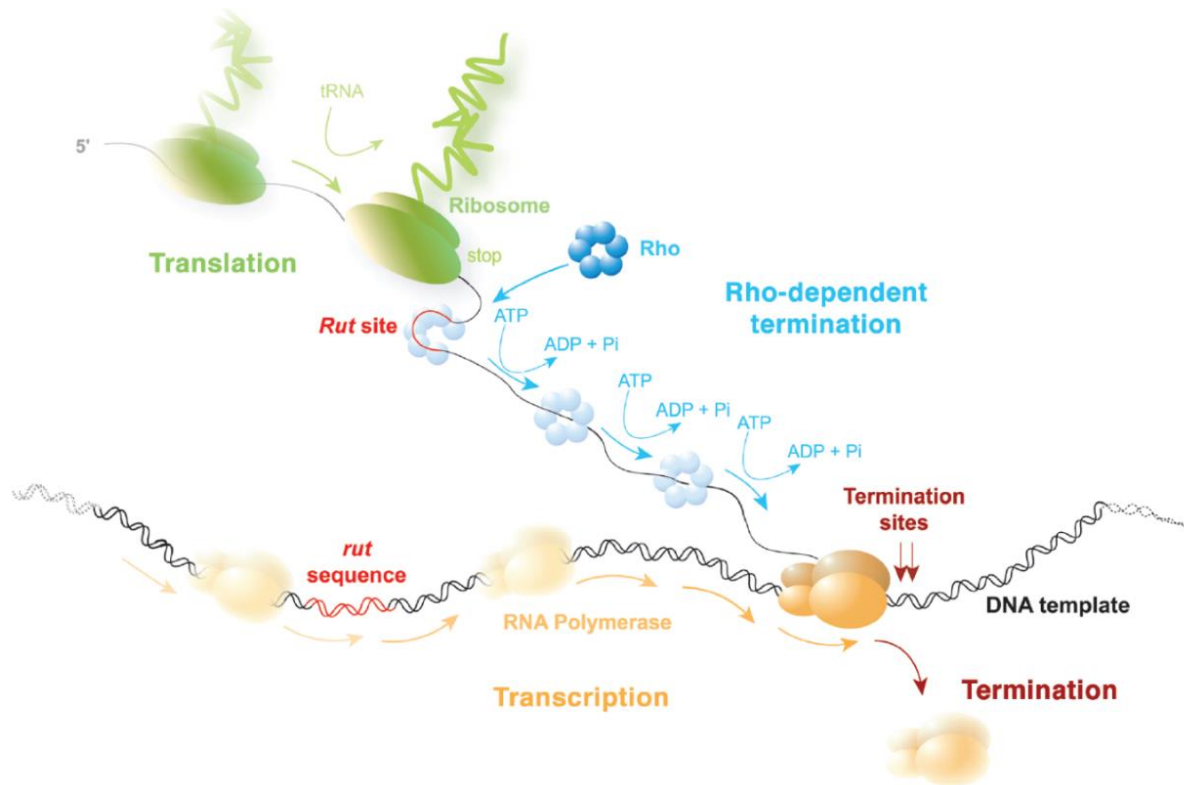


Figure 1 : Mécanisme de terminaison de la transcription Rho-dépendante. La mise à nu de l'ARN permet à Rho de se fixer sur le site Rut. Cette fixation active la capacité de Rho à hydrolyser l'ATP, puis sa translocation ATP-dépendante en direction de l'ARNP, jusqu'à provoquer la dissociation du CTE et la fin de la transcription (d'après Boudvillain et al., 2010).

Chez les bactéries, l'intervention d'ARN hélicases est souvent essentielle dans les conditions de stress, bien que secondaire dans les conditions standards de culture (Cartier *et al.*, 2010; Klostermeier, 2013; Owtrim, 2013; Redder *et al.*, 2015). Durant ma thèse je me suis consacré à l'étude du facteur Rho, une ARN hélicase initialement décrite comme capable d'induire la dissociation du complexe d'élongation de la transcription (CET) à des sites dits de terminaison Rho-dépendante (Roberts, 1969). La structure, le mécanisme d'action et les fonctions biologiques de Rho sont détaillés dans le Chapitre III. Brièvement, le facteur Rho est un homo-hexamère en forme d'anneau, qui est capable de reconnaître une séquence ARN d'interaction appelée *Rut* (*Rho utilization*) lorsque le transcrit naissant est nu (**Figure 1**). La reconnaissance de ce site *Rut* déclenche la translocation ATP-dépendante de Rho le long du brin ARN en direction de l'ARN polymérase (ARNP) (Rabhi *et al.*, 2010a). Lorsque Rho a rattrapé l'ARNP, il induit la dissociation du CET par un mécanisme qui est toujours débattu aujourd'hui (Boudvillain *et al.*, 2013; Peters *et al.*, 2011). Depuis sa découverte en 1969, les études consacrées au facteur Rho ont démontré que, chez *Escherichia coli*, cette enzyme (EcRho) participe à une large proportion (20 % à 50 %) des événements de terminaison de la transcription, contribuant ainsi à la régulation de l'expression génique, mais aussi que sa

fonction peut être étendue au maintien de l'intégrité génomique (pour revue : Boudvillain *et al.*, 2013; Gowrishankar *et al.*, 2013; Peters *et al.*, 2011). Ces éléments contribuent à expliquer pourquoi, chez *E. coli* et les espèces apparentées, Rho est une enzyme essentielle.

Cependant, ce caractère essentiel n'est peut-être pas conservé au sein du règne bactérien, comme le suggère la résistance de la plupart des espèces Gram⁺ à la bicyclomycine (BCM), un antibiotique naturel inhibant spécifiquement l'activité ATPase de Rho (voir page 97) (Kohn & Widger, 2005). Au début de cette thèse, nous n'avions qu'une vague idée de la distribution, de la conservation, et de l'expression de gènes codant pour Rho au sein de la diversité bactérienne. Seules deux études relativement anciennes, basées sur quelques séquences *rho* issues de phyla différents, proposaient que Rho est une enzyme quasi-ubiquitaire (Opperman & Richardson, 1994; Washburn *et al.*, 2001). Il nous est apparu important de vérifier/consolider cette hypothèse en profitant de la quantité considérable d'informations génomiques accumulées depuis cette époque. De plus, le peu de caractérisations biochimiques de facteurs Rho phylo-divergents ne nous permet pas d'affirmer que les mécanismes moléculaires mis en avant pour EcRho sont conservés au sein du règne bactérien. Chez certaines espèces bactériennes faisant partie du phylum des Actinobactéries, le facteur Rho possède une large insertion dans le domaine N-terminal (NTD), ce qui semble affecter son mécanisme, notamment au niveau de la reconnaissance des sites *Rut* (Nowatzke *et al.*, 1997a). Enfin, ce qui constitue un site *Rut* reste encore mal défini chez *E. coli* (Ciampi, 2006) et inconnu chez de nombreuses autres espèces bactériennes, ne nous permettant pas d'évaluer *a priori* la conservation des sites d'action (cibles) de Rho au sein de la diversité bactérienne.

Les travaux menés pendant ma thèse ont abordé deux questions principales: **i) Rho est-elle une enzyme ubiquitaire? ii) le mécanisme d'action de Rho est-il conservé ?** Dans un premier temps, en analysant les différentes bases de données génomiques et protéomiques, j'ai évalué *in-silico* la distribution et la conservation du facteur Rho au sein de la biodiversité. Cette étude présentée sous forme d'article à la page 109 du Chapitre IV a permis de confirmer que Rho est une enzyme quasi-ubiquitaire au sein de la diversité bactérienne (plus de 92 % des espèces bactériennes séquencées possèdent un gène codant pour Rho) et de montrer que des facteurs Rho issus d'autres phyla que les Actinobactéries possèdent également des insertions dans leur domaine NTD. De plus, nous avons relevé divers indices suggérant que la présence du gène *rho* est corrélée au style de vie de la bactérie et à son besoin de réguler l'expression de son génome de manière conditionnelle. Dans ce contexte, j'ai participé à la

caractérisation d'une nouvelle cible du facteur Rho située dans la partie 5' non traduite de l'opéron *pgaABCD* d'*E. coli*. Cette étude présentée page 151, a permis de mettre en avant un nouveau mécanisme de régulation génique où l'accessibilité du site *Rut* pour Rho est conditionnée par l'action de CsrA, un facteur de régulation se liant à l'ARN. Enfin, je me suis attaché à la caractérisation des activités enzymatiques de facteurs Rho possédant un domaine d'insertion dans leur NTD et plus particulièrement à celui de l'espèce *Mycobacterium tuberculosis* (MtbRho). Il a été proposé par d'autres que MtbRho est capable de provoquer la terminaison de la transcription par un mécanisme moléculaire inédit et indépendant de l'hydrolyse de NTP, expliquant l'inefficacité de la BCM à inhiber cette enzyme (Kalarickal *et al.*, 2010). Notre étude présentée en page 131 invalide cette hypothèse : bien que MtbRho présente quelques caractéristiques idiosyncratiques, son mécanisme de terminaison de la transcription reste basé sur sa capacité à hydrolyser l'ATP et à fonctionner comme une machine moléculaire se déplaçant de manière unidirectionnelle le long du brin d'ARN (**Figure 1**).

Ces résultats sont discutés dans le contexte actuel de l'étude des mécanismes de la terminaison Rho-dépendante au sein du règne bactérien (page 171). Enfin, je conclus ce manuscrit par l'évocation des perspectives offertes par mes travaux (page 176).

Chapitre I : Le métabolisme de l'ARN

La capacité d'un organisme à s'adapter aux différentes contraintes auxquelles il est soumis provient en grande partie de sa faculté à réguler l'expression de son génome. Pour certaines espèces bactériennes, cette aptitude est vitale tant ces espèces doivent naviguer entre différents écosystèmes et/ou sont soumises à des stress variés. Le métabolisme de l'ARN joue un rôle important dans ce processus d'adaptation conditionnelle. D'une part, c'est un élément fondamental de l'expression génique au travers la biosynthèse des ribosomes, la transcription et la traduction. D'autre part, la production de nombreux ARN non-codants (ARNnc) comme les ARN antisense ou les petits ARNnc encodés en *trans* (ci-après dénommés sRNA pour “*small RNAs*”) ainsi que la dégradation des différentes classes d'ARN synthétisés constituent des voies majeures de régulation (Barria *et al.*, 2013; Caldelari *et al.*, 2013; Durand *et al.*, 2015). Ces processus, bien que présents chez toutes les bactéries, ne se déroulent pas toujours de la même manière et impliquent l'action de cofacteurs différents selon les espèces (Chang *et al.*, 2006; Condon & Putzer, 2002; Durand *et al.*, 2015; Nakagawa *et al.*, 2010; Redder *et al.*, 2015).

Dans ce chapitre, je présente succinctement les principales étapes du métabolisme ARN bactérien ainsi que les différences que l'on peut rencontrer selon les espèces. J'insiste tout particulièrement sur les mécanismes impliquant le facteur Rho, le sujet de ce mémoire.

1. LA TRANSCRIPTION

La transcription représente la première étape du métabolisme ARN et du processus d'expression des gènes. Ce processus dépend de la formation du complexe d'élongation de la transcription (CET ; **Figure 2**) dont l'élément central est l'ARN polymérase (ARNP). Au cours des différentes étapes de la transcription, cette enzyme responsable de la synthèse d'ARN à partir du brin ADN matriciel subira différents réarrangements via l'action de cofacteurs. Ces réarrangements servent de points de contrôle permettant à la cellule d'assurer une expression correcte de ses gènes et le maintien de son homéostasie (Dangkulwanich *et al.*, 2014).

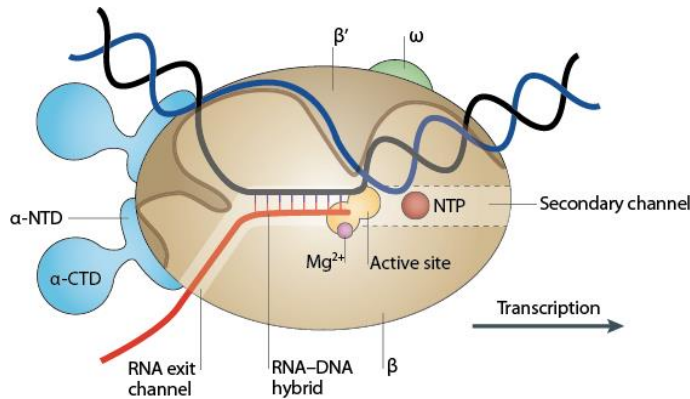


Figure 2 : Complexe d'élongation de la transcription (CET).

Le CET est composé de l'ARNP (constituée des sous-unités α en bleu clair, β et β' en marron, et ω en vert), du brin ADN (brin transcrit en noir et brin codant en bleu foncé), et du transcrit ARN (en rouge). La "bulle" de transcription est déterminée par les 13-14 pb d'ADN fondues au sein de l'ARNP et comprend également un duplexe hybride de 8-10 pb entre le transcrit naissant et le brin ADN codant. Le brin ADN non-codant est, lui, exposé à l'extérieur de l'ARNP. L'élongation du transcrit naissant se fait par ajout du NTP entrant par le canal secondaire à l'extrémité 3' de l'ARN dans le site actif de l'ARNP (cercles jaunes représentant les sous-sites de pré- et post-translocation) (d'après Santangelo & Artsimovitch, 2011)

A. L'initiation de la transcription

L'ARNP est constituée de 5 sous-unités : 2 sous-unités α , 2 sous-unités β et β' et une sous-unité ω (**Figure 2**). *In vitro* cette enzyme est capable de se fixer à l'ADN double brin et d'initier la synthèse d'ARN de manière non spécifique (Dangkulwanich *et al.*, 2014). Mais *in vivo* la transcription a besoin de démarrer à des points précis du génome appelés promoteurs (**Figure 3.A**). Pour cela, l'ARNP forme une holoenzyme avec un (co)facteur σ . Il existe plusieurs facteurs σ et, donc, holoenzymes capables de reconnaître des motifs/séquences ADN spécifiques dans le promoteur (motifs -10 et -35 voir **Figure 3.C**) et de s'y fixer. Les facteurs σ peuvent être classés en plusieurs groupes selon leurs similarités de séquences et de structures (**Figure 3.B**). Les facteurs σ du groupe 1 (par exemple, σ^{70}) servent à initier la transcription de gènes de ménage exprimés de manière continue dans la cellule (Feklístov *et al.*, 2014). Les facteurs σ des autres groupes servent à l'expression de répertoires de gènes requis pour l'adaptation à certaines conditions (par exemple, σ^{54} pour l'adaptation au froid chez les Firmicutes) (Raimann *et al.*, 2009; Wiegeshoff *et al.*, 2006) ou pour la synthèse d'organelles (σ^{28} pour la synthèse de flagelles chez *E. coli*) (Liu & Matsumura, 1995). Le nombre de facteurs σ différents codés par le génome des bactéries varie selon les espèces et leurs niches écologiques. Par exemple, les *Mycoplasma* (phylum des Ténérécutes) sont un genre regroupant majoritairement des parasites intracellulaires. Leurs niches écologiques sont

domaine 1.1 au sein de l'ARNP (**Figure 3.C**) afin d'éviter l'entrée non spécifique d'ADN dans le site actif. Le domaine 1.1 est ensuite extrudé lors de la transition vers la forme RPo (**Figure 3.C**).

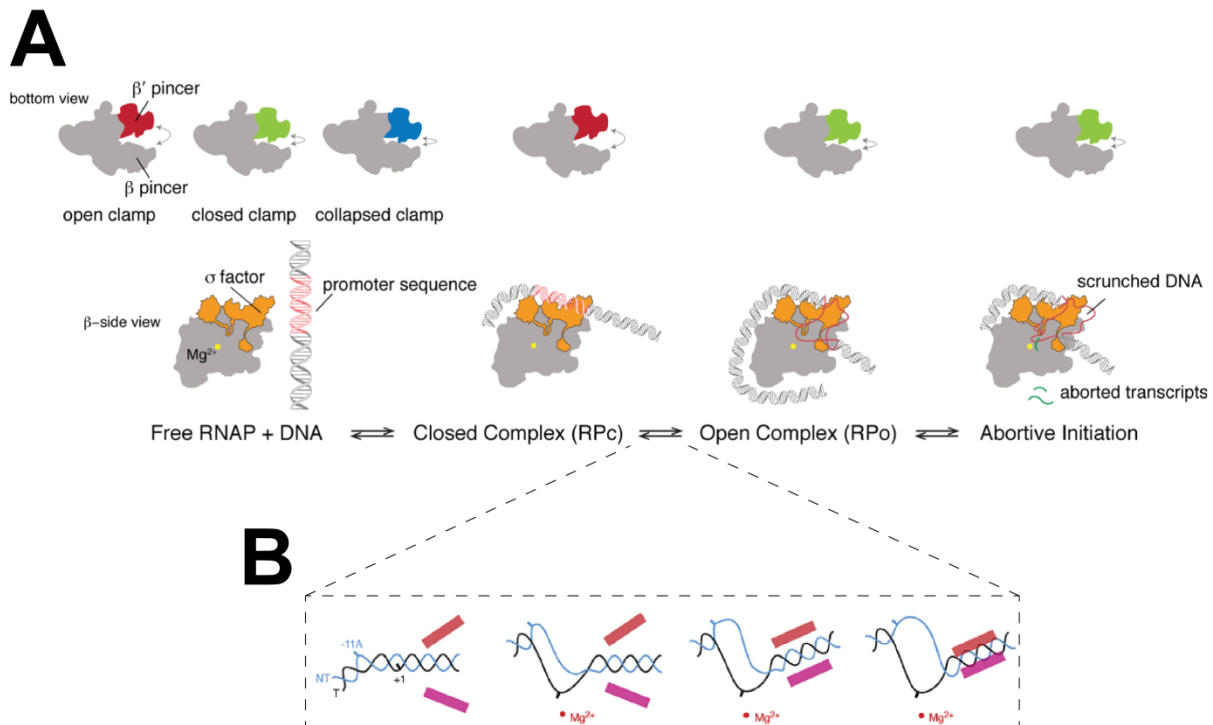


Figure 4 : Étapes caractéristiques de l'initiation de la transcription bactérienne. (A) La reconnaissance d'une séquence promotrice (rouge) par l'ARNP (gris) via le facteur σ (orange) induit la formation du complexe fermé RPc. Une succession de sous-étapes schématisées dans le panneau (B) amène à la formation du complexe ouvert RPo : torsion de l'ADN, ouverture de la bulle de transcription et contact entre la double hélice et les pinces de l'ARNP (rouge et mauve). Une fois la bulle de transcription formée dans le site actif, l'ARNP initie la synthèse d'ARN dans une phase dite de transcription abortive due à la présence du facteur σ dans le canal de sortie du transcrit ARN. Les différentes conformations observées pour la pince β/β' sont montrées en haut de chaque étape (d'après Dangkulwanich et al., 2014; Saecker et al., 2011).

Une fois le complexe RPo formé, l'ARNP commence à synthétiser une chaîne d'ARN qui, jusqu'à une longueur d'environ 11 nucléotides (nt), peut être très facilement relarguée. Ce phénomène dit de "transcription abortive" (**Figure 4.A**) résulte de l'occupation du canal de sortie de l'ARN par le domaine 3.2 du facteur σ (**Figure 3.B**), aussi appelé "doigt σ ", qui s'oppose à l'élongation du transcrit (Murakami et al., 2002). Néanmoins, il arrive que ce domaine soit déplacé lors de l'élongation du transcrit ARN, un phénomène accompagné par le mouvement de l'enzyme hors de la zone du promoteur et son entrée dans la phase d'élongation de la transcription. A partir de ce moment, les multiples contacts entre l'ARNP, l'ADN et l'ARN assurent une grande stabilité au CET (Peters et al., 2011).

B. L'élongation de la transcription

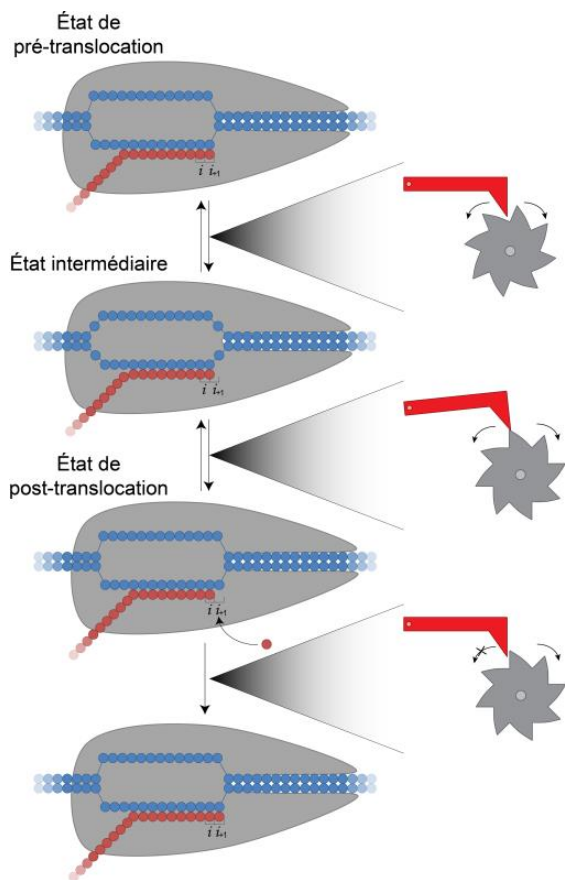


Figure 5 : Translocation du CTE selon le modèle Brownian Ratchet.

La translocation de l'ARNP implique trois étapes. La fixation et le branchement du NTP au site $i+1$ est une étape irréversible, favorisant la translocation de l'holoenzyme vers l'avant.

Durant l'élongation, l'addition de ribonucléotides triphosphate (NTP) au transcrit ARN provoque la translocation de l'ARNP le long de la double hélice ADN. Cette translocation se fait une paire de base à la fois et à des vitesses moyennes pouvant varier significativement (10-90 pb/s) suivant le gène transcrit et/ou les conditions expérimentales (Abbondanzieri *et al.*, 2005; Dennis *et al.*, 2009; Proshkin *et al.*, 2010). Il a été démontré que, pour assurer cette translocation, l'ARNP utilise un mécanisme dit de *Brownian ratchet* (**Figure 5**) (Abbondanzieri *et al.*, 2005; Dangkulwanich *et al.*, 2014; Guajardo & Sousa, 1997).

Différents signaux et facteurs affectent ou contribuent à réguler la phase d'élongation de la transcription (**Figure 6**). Ces éléments sont discutés brièvement ci-dessous.

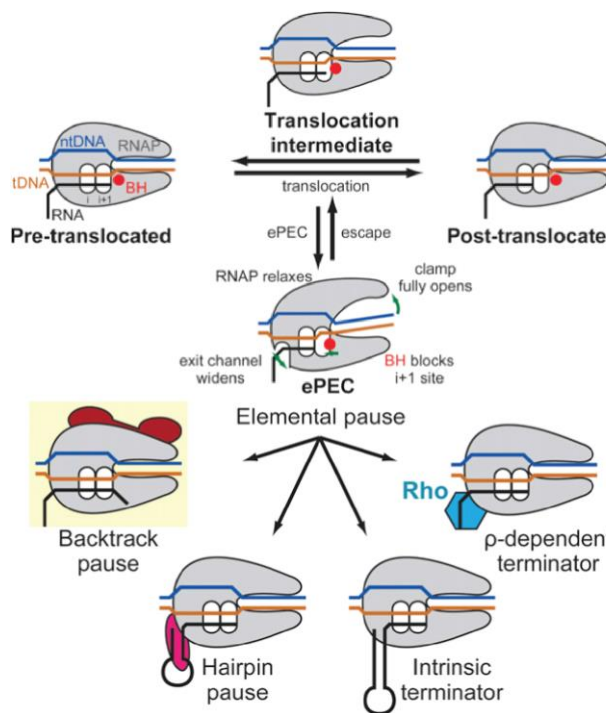


Figure 6 : Processus élémentaire d'élongation de la transcription

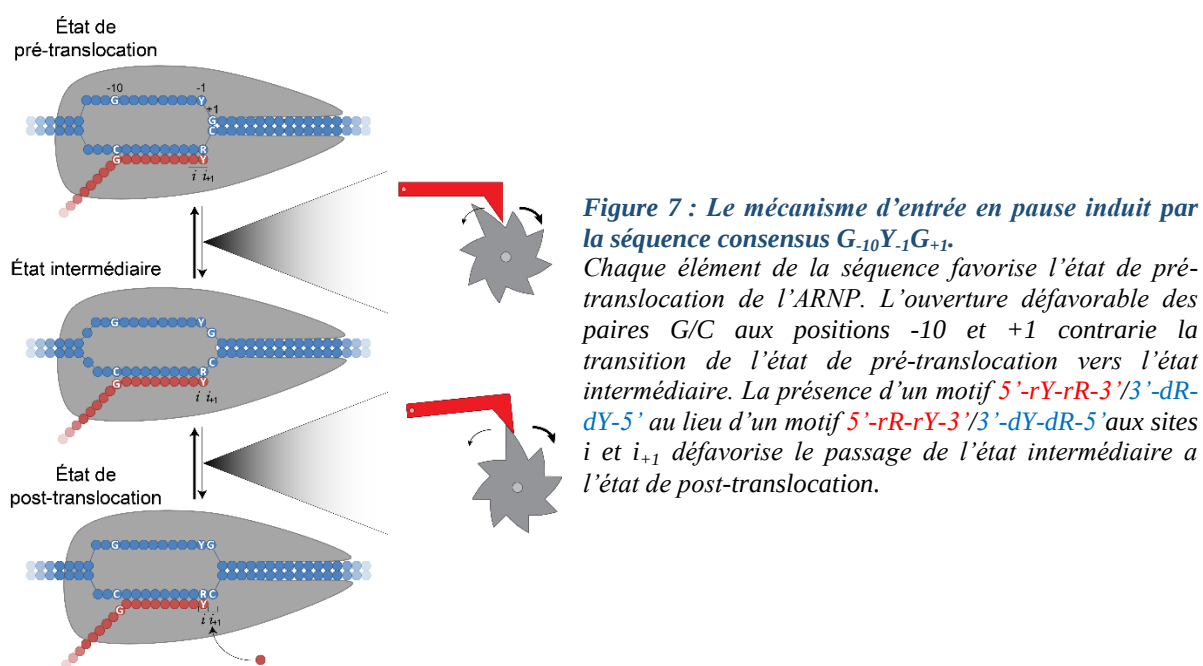
Durant chaque étape d'élongation, l'ARNP transite entre trois configurations lui permettant la translocation et l'ajout d'un nucléotide à l'extrémité de la chaîne ARN. Dans la configuration intermédiaire ("translocation intermédiaire" sur la figure), l'ARNP peut échouer à réengager ses contacts avec l'ADN, ce qui la conduit à entrer dans un état de pause élémentaire (ePEC). Cet état est associé à différents signaux régulateurs de l'élongation : 1) signal de pause prolongée qui est induite par la formation d'une structure en tige-boucle dans l'ARN (classe I) ou par un phénomène de rétro-translocation ("backtracking") de l'ARNP (classe II) et qui est régulé par les facteurs NusA (rose) ou NusG (rouge) ; 2) signal de terminaison induit par un terminateur intrinsèque (qui comprend un motif tige-boucle au sein de l'ARN plus stable que pour la pause de classe I) ou par l'action du facteur Rho (bleu). La pause de classe II (carré jaune) peut conduire à un "backtracking" encore plus prononcé et à l'arrêt complet du CET. Les CET arrêtés nécessitent des facteurs auxiliaires pour leur réamorçage et sont néfastes pour l'intégrité génomique (d'après Weixlbaumer et al., 2013).

a. La pause

L'élongation n'est pas un processus continu. L'ARNP va répétitivement adopter un état transitoire dans lequel elle rompt ses contacts avec la double hélice ADN afin de pouvoir diffuser le long de celle-ci (**Figure 6**). Dans la plupart des cas, ces contacts seront rétablis un nucléotide en aval, mais régulièrement l'enzyme échoue à les rétablir et entraîne le CET dans un état dit de pause élémentaire (ePEC sur la **Figure 6**) (Weixlbaumer et al., 2013). En 2014, deux études ont permis d'identifier une séquence consensus susceptible d'induire la pause élémentaire de l'ARNP d'*E. coli* (Larson et al., 2014; Vvedenskaya et al., 2014). Cette séquence consensus se compose d'un G en position -10, d'un Y (C ou T) à la position -1 et d'un G à la position +1 du brin ADN non transcrit. Chacun de ces résidus est prédit pour favoriser l'état de pré-translocation de l'enzyme et défavoriser l'état de post-translocation (**Figure 6**). Pour assurer sa translocation, l'ARNP doit provoquer la séparation de la double hélice ADN à la position +1. Cette étape est défavorisée par la présence d'une paire de base dG/dC plutôt que dA/dT. Un même cas de figure se présente à la position -10 avec la séparation de la paire de base rG/dC du duplexe ARN:ADN (**Figure 7**). De plus, le site actif

de l'ARNP interagit préférentiellement avec les motifs de type 5'-rR-rY-3'/3'-dY-dR-5' (Hein *et al.*, 2011). Les présences d'une pyrimidine (Y) et d'une guanine aux positions -1 et +1, respectivement, vont donc défavoriser le passage de l'enzyme à l'état de post-translocation qui devra accueillir ultérieurement un motif 5'-rY-rG-3'/3'-dR-dC-5' dans le site actif (**Figure 7**).

La séquence consensus "G₋₁₀Y₋₁G₊₁" est retrouvée près de 20 000 fois dans le génome d'*E. coli* et plus particulièrement près du *rbs* (*Ribosome Binding Site*) contenant la séquence Shine Dalgarno (SD - voir page 39) (Larson *et al.*, 2014). L'hypothèse des auteurs est que la pause de l'ARNP à ces sites favorise le couplage entre la transcription et la traduction (Larson *et al.*, 2014). Cette séquence consensus peut subir de légères variations selon les espèces mais reste néanmoins reconnue par différentes ARNPs bactériennes, voire même par l'ARNP II de *Saccharomyces cerevisiae*, démontrant que ce phénomène de pause élémentaire (**Figure 6** & **Figure 7**) est très conservé au sein de l'arbre de vie (Larson *et al.*, 2014).



La pause élémentaire dure rarement plus de quelques secondes avant que l'ARNP ne réamorçe la translocation (Neuman *et al.*, 2003). Mais elle peut aussi conduire à des états de pauses plus longs (**Figure 6**) suivant le contexte de séquence local (Artsimovitch & Landick, 2000; Neuman *et al.*, 2003; Weixlbaumer *et al.*, 2013). Il existe deux types majeurs de pauses longues, chacun impliquant des modifications structurales spécifiques qui stabilisent le CET dans un état non-productif (Artsimovitch & Landick, 2000).

La pause de classe I est induite par la formation dans le transcrit ARN d'une structure de type tige-boucle ("*hairpin*") à l'extrémité du canal de sortie de l'ARN (**Figure 6**). La formation de cette structure maintient l'extrémité 3'-OH du brin ARN incorrectement alignée dans le site actif (Artsimovitch & Landick, 1998; Chan *et al.*, 1997) par un mécanisme allostérique impliquant des contacts directs entre la tige-boucle et l'ARNP (Hein *et al.*, 2014; Touloukhonov *et al.*, 2001; Touloukhonov & Landick, 2003).

La pause de classe II se produit quand l'ARNP effectue un mouvement de recul de quelques nucléotides (jusqu'à une dizaine environ) le long de l'ADN. Ce mouvement de rétro-translocation, communément appelé "*backtracking*", provoque l'envahissement par l'extrémité 3' de l'ARN du canal d'entrée des NTP (dit "canal secondaire" de l'ARNP) qui est normalement utilisé par ces-derniers pour accéder au site actif de l'ARNP (**Figure 6 & Figure 8**). Contrairement à la pause de classe I qui est systématiquement réversible, le mouvement de rétro-translocation de l'ARNP peut s'amplifier et conduire à la formation d'un CET arrêté (**Figure 6 & Figure 8**) incapable, seul, de réamorcer l'élongation de la transcription. Les CET arrêtés jouent un rôle néfaste pour l'organisme car ils peuvent être à l'origine de R-Loops compromettant l'intégrité du génome (Dutta *et al.*, 2011). Les R-Loops sont des duplexes ARN:ADN qui peuvent résulter de la collision du complexe de réplication avec une ARNP arrêtée (**Figure 8**). Ces hybrides sont susceptibles d'être reconnus comme des fragments d'Okazaki et de provoquer l'initiation de la réplication causant alors une coupure simple brin dans l'ADN (**Figure 8**) (Dutta *et al.*, 2011). Deux facteurs ubiquitaires, GreA et GreB, (Opalka *et al.*, 2003) contribuent à limiter la formation de CET arrêtés (Borukhov *et al.*, 1993). Chez *E. coli*, ces deux protéines paralogues sont capables de se fixer près du canal secondaire de l'ARNP (**Figure 2**) et d'y provoquer la coupure de la partie 3'-extrudée du transcrit ARN (**Figure 8**). GreB est capable de provoquer la coupure d'ARN extrudé d'une longueur pouvant aller jusqu'à environ 18 nucléotides, ce qui permet à l'ARNP de reprendre l'élongation, alors que GreA ne provoque la coupure que de 2 à 3 nucléotides et serait plus impliquée dans la prévention des phases précoces de rétro-translocation (Borukhov *et al.*, 2005 ; Rutherford *et al.*, 2007). Un autre élément permettant d'éviter la rétro-translocation de l'ARNP est le couplage entre la traduction et la transcription. *In vivo*, lors de la transcription d'une séquence codante, il y a contact entre le ribosome "*leader*" du polysome et l'ARNP (voir page 41), ce qui pousse cette dernière vers l'avant, et empêche sa rétro-translocation (Proshkin *et al.*, 2010). Dans des conditions de défaut en nutriments ("*starvation*") affectant le couplage entre traduction et transcription, la protéine DksA (plus

connue pour son rôle régulateur de l'initiation) est capable elle aussi de prévenir la formation de complexes arrêtés (Tehranchi *et al.*, 2010). Cependant *dksA* n'est retrouvée que dans le génome des Protéobactéries et semble absente des autres phyla. Une alternative à DksA est le facteur Rho, sujet de cette thèse, qui a été décrit comme inhibant la propagation de R-Loops dans le génome (Harinarayanan & Gowrishankar, 2003; Leela *et al.*, 2013). Le rôle du facteur Rho dans la répression des R-Loops est détaillé dans le Chapitre III de ce mémoire (voir page 94).

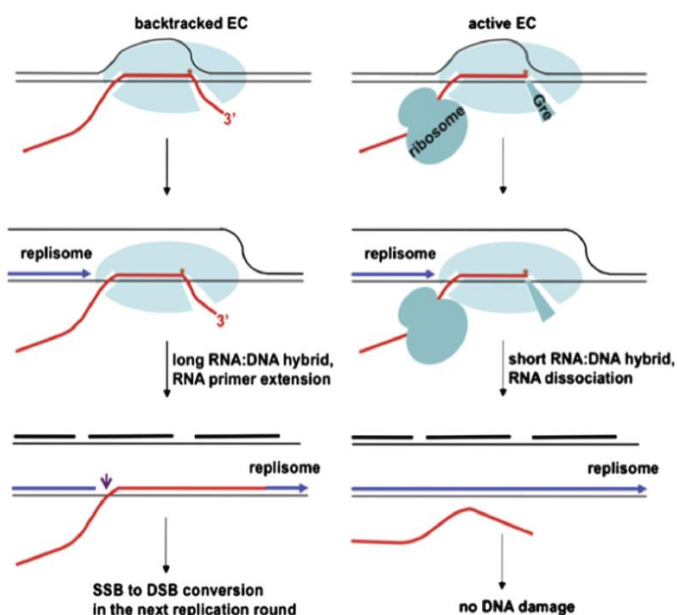


Figure 8 La formation de R-Loops induite par les complexes arrêtés.

Quand l'ARNP est arrêtée via un phénomène de rétro-translocation, la collision avec le réplisome induit la dissociation du CET (EC, sur la figure). L'extrémité 3' de l'ARN extrudée dans le canal d'entrée du NTP pourra alors s'hybrider avec le brin ADN et créer un nouveau point d'initiation de la réplication. Ce phénomène entraîne la création d'une cassure simple brin dans l'ADN (indiquée par la flèche violette) voire une cassure double brin au prochain cycle de réplication. Ce phénomène est limité par l'action des facteurs Gre, et du couplage entre la réplication et la traduction (d'après Dutta *et al.*, 2011).

Il existe également un autre type de pause, localisée près du promoteur et qui est induite par la présence du facteur σ . Au début de la phase d'élongation, les domaines 1.1 et 3.2 du facteur σ sont exclus de leurs sites d'interaction avec l'ARNP mais le domaine 2 reste fixé, ce qui rend le facteur σ susceptible de reconnaître des séquences de type "promoteur" (Figure 3) et de s'y associer (Mooney *et al.*, 2005). Cette fixation provoque l'arrêt temporaire de l'ARNP jusqu'au décrochage complet du facteur σ (Mooney *et al.*, 2009a). Les séquences susceptibles d'induire ce type de pause sont généralement retrouvées à proximité des promoteurs et seraient impliquées dans le phénomène dit de pause "proximale", utile pour coordonner le couplage entre la transcription et la traduction (Mooney *et al.*, 2005).

b. La régulation de l'élongation

La pause est un événement important permettant le bon déroulement du processus de transcription (Landick, 2006) mais qui, dans le cas de la pause de classe II, peut avoir des conséquences néfastes pour la bactérie. Il est donc logique que ces événements soient

étroitement régulés par de nombreux facteurs dont certains ont déjà été évoqués plus haut. Parmi les facteurs majeurs, NusA et NusG sont retrouvés chez quasiment toutes les bactéries, à l'exception de certains parasites intracellulaires (Belogurov *et al.*, 2007) et participent activement à la régulation de la transcription notamment en modulant les signaux de pause.

NusA fut initialement décrit comme faisant partie du complexe d'anti-termination de la transcription de l'opéron *rrn* (voir page 34) (Squires *et al.*, 1993). Ce facteur, essentiel chez *E. coli* et *B. subtilis* (Ingham *et al.*, 1999; Zheng & Friedman, 1994), est capable de diminuer la vitesse globale d'élongation en favorisant la pause de classe I (**Figure 6**) et de stimuler la terminaison intrinsèque (voir page 28). Il est proposé que NusA forme de multiples contacts avec l'ARNP près du canal de sortie de l'ARN dès que l'ADN ne contacte plus la sous-unité α et que le facteur σ se dissocie (Mooney *et al.*, 2009a; Washburn & Gottesman, 2015), c'est à dire lorsque le complexe transcriptionnel entre dans la phase d'élongation (**Figure 9**). De plus, le NTD de NusA est capable d'interagir avec le transcrit ARN et de stabiliser ainsi la formation de la structure tige-boucle, entraînant la pause de l'ARNP (**Figure 6**) (Kolb *et al.*, 2014). Les caractéristiques de NusA sont comparables chez *E. coli* et *B. subtilis* (Ingham *et al.*, 1999; Yang *et al.*, 2009; Zheng & Friedman, 1994) mais il peut exister des différences notables chez d'autres espèces. Par exemple, le facteur NusA de *Mycobacterium bovis* ne semble pas capable de stimuler la terminaison intrinsèque (Czyz *et al.*, 2014).

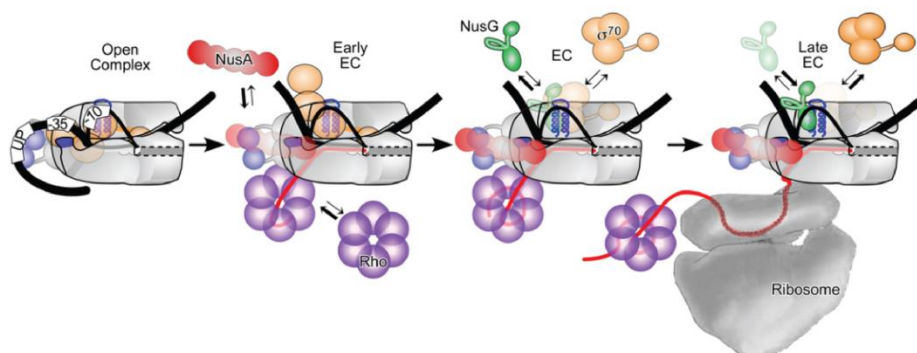


Figure 9 : Modèle d'échange de facteurs de régulation entre l'initiation et l'élongation. Le passage d'un état *RPc* à un état *RPO* provoque la libération du CTD de la sous-unité α de l'ARNP, ce qui permet le recrutement de NusA. Par la suite, NusG provoque le relargage du facteur σ et assure le couplage physique avec le premier ribosome du polysome en se fixant également à la protéine S10 (non représenté ici) (d'après Mooney *et al.*, 2009a).

A l'inverse de NusA, NusG est capable d'augmenter la vitesse globale de l'élongation de la transcription et ce à plusieurs échelons. Le NTD de NusG forme des contacts avec l'ARNP à des sites recouvrant ceux du facteur σ (Belogurov *et al.*, 2007). La compétition entre ces deux facteurs pourrait favoriser le relargage du facteur σ (**Figure 9**) (Mooney *et al.*,

2005). Cette hypothèse est appuyée par le fait que, *in vivo*, l'association entre NusG et l'ARNP n'est détectable uniquement que quand l'association avec le facteur σ est rompue (Mooney *et al.*, 2009a). Cette dissociation pourrait avoir lieu durant la pause σ -dépendante et est prédite pour être un checkpoint assurant le couplage entre la transcription et la traduction (Mooney *et al.*, 2005). Ce couplage est garanti physiquement par NusG dont le CTD peut se fixer à la protéine ribosomique NusE, alias S10 (voir page 41) (Burmann *et al.*, 2010). NusG sert alors de pont entre l'ARNP et le ribosome “*leader*” du polysome, de ce fait inhibant la rétro-translocation de l'ARNP ainsi que la terminaison précoce de la transcription par le facteur Rho (voir page 30 et Chapitre III) (pour revue : Boudvillain *et al.*, 2013; Washburn & Gottesman, 2015). On peut noter également que la fixation de NusG sur l'ARNP durant l'élongation stabilise les interactions de l'ARNP avec l'ADN, ce qui inhibe également la rétro-translocation (indépendamment de la présence du ribosome) et rend le CET plus processif (Herbert *et al.*, 2010; Svetlov & Nudler, 2011).

Prises ensemble, ces données montrent que NusG est un facteur primordial assurant l'entrée dans la phase d'élongation de la transcription, et favorisant son déroulement. Cette importance est telle que la présence et la fonction de NusG semblent avoir été conservées dans les trois règnes du vivant au cours de l'évolution (Blombach *et al.*, 2013; Yakhnin & Babitzke, 2014). Cependant, NusG pourrait posséder d'autres rôles ou des rôles différents chez certaines bactéries. De façon intrigante, par exemple, le facteur NusG de *M. bovis* stimule la terminaison intrinsèque *in vitro* alors que celui d'*E. coli* l'atténue (Czyz *et al.*, 2014). Chez *Bacillus subtilis*, NusG induit la pause de l'ARNP à deux sites régulateurs situés dans la partie 5' non-traduite de l'opéron *trp* (Yakhnin *et al.*, 2008). Ces pauses seraient liées à la reconnaissance par NusG de séquences riches en T dans le brin non-matriciel exposées à la surface de l'ARNP. Cette reconnaissance impliquerait de courts motifs situés dans le NTD de NusG de *B. subtilis* qui sont absents du facteur NusG d'*E. coli* (Yakhnin *et al.*, 2008). Un mécanisme analogue a cependant été identifié chez *E. coli* pour la protéine paralogue RfaH qui possède un CTD similaire à NusG mais un NTD différent (Yakhnin & Babitzke, 2014). La présence de *rfaH* dans le génome apparaît être exclusive aux Protéobactéries. A la différence de NusG, l'action de RfaH semble être restreinte à des gènes issus de transferts horizontaux (plasmides ou îlots génomiques) et contenant un site *ops* à proximité du promoteur (Tomar & Artsimovitch, 2013). La séquence de ces sites *ops* exposée à la surface de l'ARNP permet le recrutement de RfaH et sa fixation sur l'ARNP. Le CTD de RfaH, tout

comme celui de NusG, va alors permettre le couplage entre la transcription et la traduction en se fixant à la protéine ribosomique NusE/S10 (Tomar & Artsimovitch, 2013).

C. La terminaison de la transcription

Plusieurs voies peuvent conduire à l'interruption de la transcription. Cette interruption peut être irréversible suite à la dissociation du CET induite par un signal cellulaire (on parle alors de terminaison de la transcription) ou simplement refléter un état inactif du CET qui ne peut alors retrouver tout seul un registre catalytiquement-compétent (on parle alors simplement d'arrêt de la transcription). Comme indiqué précédemment, ce dernier cas se présente lors d'une rétro-translocation trop importante de l'ARNP (Dutta *et al.*, 2011) mais il peut également être occasionné par la présence de dommages (par exemple, une cassure simple brin) dans l'ADN (Howan *et al.*, 2014; Monnet *et al.*, 2013). La terminaison de la transcription, quant à elle, est induite par deux types de mécanismes distincts qui sont brièvement décrits ci-dessous:

- La terminaison intrinsèque (ou Rho-indépendante) qui dépend de la présence dans l'ADN d'une séquence codant pour un motif ARN spécifique.
- La terminaison facteur-dépendante qui requière l'action d'un facteur protéique. Deux facteurs de terminaison majeurs sont connus : Mfd impliqué dans la voie TCR (*Transcription Coupled Repair*) (Monnet *et al.*, 2013) et Rho dont les caractéristiques sont détaillées au Chapitre III.

a. La terminaison intrinsèque

La terminaison intrinsèque intervient principalement à la fin des opérons où elle gouverne la formation des extrémités 3' des ARN messagers (ARNm) (Nicolas *et al.*, 2012; Peters *et al.*, 2011). L'analyse globale des transcriptomes de *Mycoplasma pneumoniae*, *Synechococcus elongatus* et *B. subtilis* suggère que cette configuration est très largement majoritaire chez ces espèces (Guell *et al.*, 2009; Nicolas *et al.*, 2012; Vijayan *et al.*, 2011). En revanche, des expériences comparables ne retrouvent pas de motifs caractéristiques de la terminaison intrinsèque (voir ci-dessous) pour des proportions significatives (>20%) de transcrits chez *Helicobacter pylori* ou *E. coli* (Peters *et al.*, 2009; Peters *et al.*, 2011; Sharma *et al.*, 2010). Diverses analyses bio-informatiques de la distribution de motifs de terminaison intrinsèque au sein des génomes bactériens suggèrent que cette voie de terminaison de la

transcription n'est pas prépondérante chez certaines espèces comme *M. tuberculosis* (Ermolaeva *et al.*, 2000; Mitra *et al.*, 2009; Washio *et al.*, 1998). Cette information est néanmoins à prendre avec prudence dans la mesure où des terminateurs "imparfaits" (par exemple, contenant des appariements non-canoniques) peuvent facilement échapper à une détection *in silico* (Washio *et al.*, 1998), notamment lors de l'analyse de génomes riches en G/C (Brenneis *et al.*, 2007).

Il est généralement admis que la terminaison intrinsèque intervient quand l'ARNP rencontre une séquence codant pour un motif ARN composé i) d'une tige-boucle dont la tige est d'environ 8 pb et riche en paires G/C et ii) d'une séquence 3' adjacente riche en résidus uraciles (avec généralement une succession de U juste après la tige-boucle ; (Peters *et al.*, 2011). Durant la transcription de ce motif, la séquence riche en résidus uraciles forme avec le brin matrice un hybride ARN:ADN peu stable dans le site actif, ce qui favorise la pause de la polymérase (Gusarov & Nudler, 1999). Cette pause, à son tour, favorise la formation de la structure en tige-boucle dans le canal de sortie de l'ARN au sein de l'ARNP (Peters *et al.*, 2011)). Cet événement conduit à la déstabilisation et à la dissociation du CET suivant un mécanisme qui reste aujourd'hui débattu (**Figure 10**).

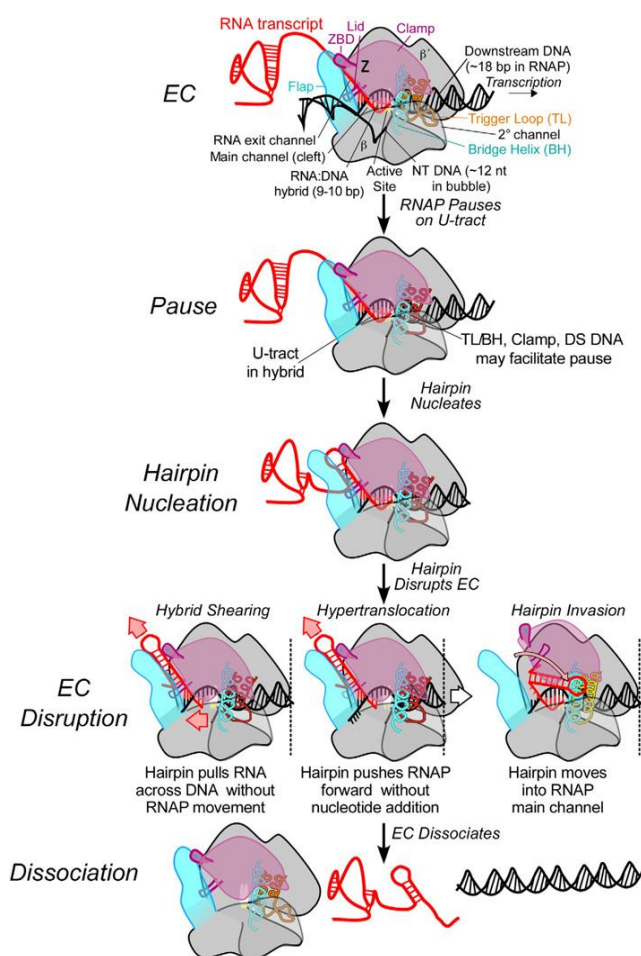


Figure 10 : Mécanisme de la terminaison intrinsèque.

La transcription de la séquence riche en U induit la pause de l'ARNP. Cette pause favorise la formation de la structure tige boucle dans le canal de sortie de l'ARN. La formation de cette structure déstabilise le CET suivant un mécanisme pour lequel trois variantes ont été proposées. La déstabilisation du CET provoque le décrochage de l'ARNP de l'ADN et le relargage du transcrit ARN (d'après Peters *et al.*, 2011).

Certains spécialistes proposent que la formation de la tige-boucle est suffisante pour détruire l'hybride ARN:ADN peu stable en commençant par les 3 pb les plus proches de la structure en tige-boucle (**Figure 10** -“*hybrid shearing*” ;(Peters *et al.*, 2011). D'autres proposent que la tige-boucle provoque une hyper-translocation de l'ARNP sans que la chaîne ARN ne soit allongée, ce qui conduit à un raccourcissement de l'hybride ARN:ADN et à un effondrement de la bulle de transcription (**Figure 10**-“*hypertranslocation*” ; (Santangelo & Roberts, 2004). Enfin, suivant les mêmes principes que pour la pause de classe I, certains auteurs défendent un modèle où l'invasion du canal de sortie de l'ARN par la tige-boucle provoque une déstabilisation allostérique du CET (**Figure 10**-“*hairpin invasion*”, encore appelé modèle allostérique ; (Epshtein *et al.*, 2007). Il est difficile de trancher avec les outils expérimentaux actuels même si des expériences de nano-manipulation à l'échelle de la molécule unique suggèrent que différents mécanismes pourraient être mis en œuvre suivant le terminateur étudié (Larson *et al.*, 2008).

Un élément clé de la terminaison intrinsèque est la coordination entre la formation de la structure secondaire et la translocation de l'ARNP. Cette coordination est aidée par l'entrée en pause de l'ARNP. L'efficacité de terminaison peut être modulée *in vitro* par l'utilisation de mutants plus ou moins rapides de l'ARNP d'*E. coli* (Jin *et al.*, 1992) et peut être améliorée par NusA qui va stabiliser la tige-boucle dans le canal de sortie (Yakhnin & Babitzke, 2002). Une étude récente démontre que l'ARNP de *M. bovis* est capable d'utiliser plus efficacement que celle d'*E. coli* des terminateurs intrinsèques dont l'extrémité est moins riche en uraciles que de coutume (Czyz *et al.*, 2014). Les auteurs expliquent cela par le fait que le CET de *M. bovis* est plus lent et, probablement, moins stable (Czyz *et al.*, 2014).

b. La terminaison facteur-dépendante.

Les facteurs de terminaison Mfd et Rho sont tous deux des enzymes possédant une activité hélicase/translocase (voir Chapitre II page 51) mais qui opèrent par des mécanismes significativement différents.

Mfd est une ADN hélicase capable de déplacer une ARNP arrêtée par un dommage dans l'ADN afin de rendre ce dernier accessible à la machinerie de réparation (Howan *et al.*, 2012; Monnet *et al.*, 2013). Au cours du processus, Mfd forme avec l'ARNP un complexe de longue durée de vie sur l'ADN, se mouvant à ~ 4bp/s (pour s'éloigner du site de dommage), mais duquel le transcrit est rapidement relargué (Graves *et al.*, 2015).

Rho est une ARN hélicase qui, lorsqu'elle peut accéder au transcrit naissant, va s'y associer et se déplacer le long de celui-ci pour rattraper l'ARNP et induire la dissociation du CET (**Figure 1**). Les caractéristiques de Rho étant détaillées dans le Chapitre III (page 69), seuls quelques éléments éclairant les autres sections du présent chapitre sont exposés ci-dessous.

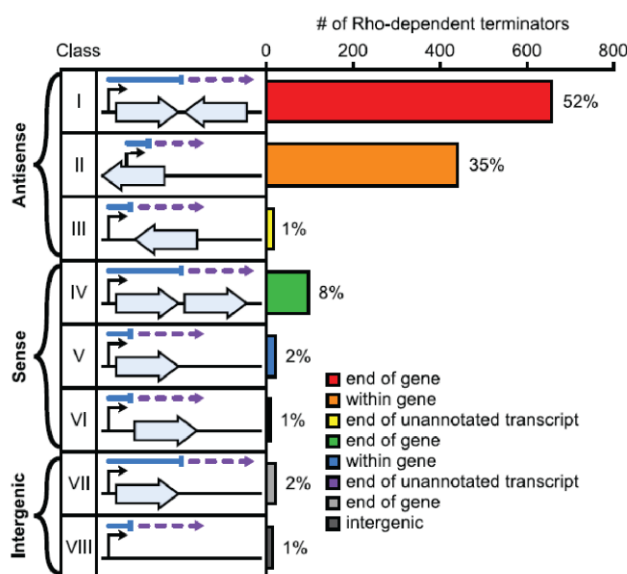


Figure 11 : Evénements de terminaison de la transcription Rho-dépendante chez *E. coli*.

Schéma illustrant l'orientation des événements de transcription et de terminaison Rho-dépendante par rapport aux gènes annotés (grosses flèches remplies en bleu clair). Les flèches noires représentent les promoteurs transcriptionnels. Les traits bleus indiquent les régions transcrites lorsque Rho est fonctionnel alors que les flèches pointillées violettes représentent les zones de transcription étendue observées lorsque Rho est inhibé par l'ajout de bicyclomycine dans le milieu de culture. Les proportions des différentes classes de terminaison Rho-dépendantes sont également indiquées (d'après (Peters et al., 2012)).

La terminaison Rho-dépendante représenterait de 20 à 50 % des événements de terminaison transcriptionnelle chez *E. coli* (Peters et al., 2009; Zhu & von Hippel, 1998). C'est une quantité difficile à estimer car de nombreux terminateurs Rho-dépendants (terminateurs intra-géniques, par exemple; voir ci-dessous) pourraient n'être fonctionnels que dans certaines conditions. Une des rares études ayant exploré systématiquement l'impact des conditions expérimentales (104 conditions testées) sur le transcriptome a été menée chez *B. subtilis* où seulement une centaine d'unités transcriptionnelles seraient assujetties à (ou réprimées par) la terminaison Rho-dépendante (Nicolas et al., 2012). Par comparaison, 1264 unités transcriptionnelles Rho-dépendantes ont été identifiées chez *E. coli* pour une seule condition de culture (Peters et al., 2012). De plus, cette analyse transcriptomique a démontré que, chez *E. coli*, la terminaison Rho-dépendante intervient aussi bien à la fin qu'à l'intérieur des unités transcriptionnelles (voir page 47), ainsi que dans les zones inter-géniques (**Figure 11**) (Peters et al., 2012). Néanmoins, la très grande majorité (88%) des sites de terminaison Rho-dépendante correspond à la répression d'événements de transcription antisense (**Figure 11**) (Peters et al., 2012). Ce rôle inattendu de Rho dans le contrôle de la transcription antisense (et, donc, du "bruit" transcriptionnel) est retrouvé chez *B. subtilis* (Nicolas et al., 2012).

Chez *E. coli*, plusieurs facteurs sont capables de moduler la terminaison de la transcription Rho-dépendante. Certains agissent en formant des contacts directement avec la protéine Rho. Parmi ceux-ci, les protéines YaeO (Gutierrez *et al.*, 2007) et Hfq (Rabhi *et al.*, 2011a), tout comme la protéine Psu du bacteriophage P4 (Pani *et al.*, 2009), ont été décrites comme étant capables de diminuer l'efficacité de terminaison Rho-dépendante. A l'inverse, l'interaction entre le facteur NusG (via son CTD) et Rho stimule souvent la terminaison de la transcription, favorisant l'utilisation de terminateurs faibles et/ou une terminaison plus précoce (à des sites plus proches du promoteur ; (Chalissery *et al.*, 2011; Mooney *et al.*, 2009b).

D'autres facteurs semblent pouvoir moduler la terminaison Rho-dépendante de manière indirecte (ou en *trans*). Par exemple, la présence de la protéine H-NS peut augmenter l'efficacité de la terminaison Rho-dépendante à certains loci du génome (Kotlajich *et al.*, 2015; Nagarajavel *et al.*, 2007; Peters *et al.*, 2012; Rajkumari & Gowrishankar, 2001; Tran *et al.*, 2011). Tout comme Rho, H-NS semble jouer un rôle important dans la répression de la transcription antisense (Singh *et al.*, 2014) et co-localise largement avec les sites de terminaison Rho-dépendante (Peters *et al.*, 2012). Le ribosome, du fait du couplage entre la transcription et la traduction (McGary & Nudler, 2013; Newton *et al.*, 1965; Yanofsky *et al.*, 1971), peut, lui, altérer la capacité de Rho à terminer la transcription en empêchant la protéine d'interagir avec le transcrit (Adhya & Gottesman, 1978; Boudvillain *et al.*, 2013; Hussein *et al.*, 2015; McGary & Nudler, 2013). Le facteur NusA est également capable de réguler la terminaison Rho-dépendante (Zheng & Friedman, 1994) mais son mode d'action reste mal compris et est probablement multifactoriel. NusA est capable d'interagir physiquement avec l'ARNP (Mooney *et al.*, 2009a; Washburn & Gottesman, 2015) ainsi que, probablement, avec Rho (Schmidt & Chamberlin, 1984) et/ou d'autres facteurs régulant l'élongation de la transcription (comme la protéine λ N) (Mishra & Sen, 2015). Ces interactions, notamment au sein de complexes dits "d'anti-terminaison" (Washburn & Gottesman, 2015) (voir page 34), et/ou leurs effets indirects sur le couplage transcription/traduction confèrent à NusA la capacité d'inhiber la terminaison Rho-dépendante (Burns *et al.*, 1998). En accord avec cette fonction inhibitrice, la délétion de *nusA* n'est viable chez *E. coli* que si l'activité de Rho est diminuée par une mutation délétère (Zheng & Friedman, 1994). Cependant, la présence de NusA semble stimuler la terminaison Rho-dépendante à certains sites du génome, probablement parce que NusA y favorise la pause de l'ARNP (Carlomagno & Nappo, 2003). De plus, il a été proposé que l'action de NusA est généralement requise pour la répression

Rho-dépendante des gènes étrangers, notamment des gènes de prophages (Cardinale *et al.*, 2008) qui représentent une cible préférentielle pour Rho (~ 11 % des évènements de terminaison Rho-dépendante chez *E. coli* ; voir page 93) (Peters *et al.*, 2009). Cette fonction de NusA a été néanmoins fortement contestée dans une étude récente du transcriptome Rho-dépendant chez *E. coli* (Peters *et al.*, 2012).

2. LA TRADUCTION

Au cours de la traduction, la cellule exprime son patrimoine génétique au travers la synthèse de protéines. Différents types d'ARN sont impliqués dans ce processus. Les ARN ribosomiques (ARNr) associés aux protéines ribosomiques forment la machinerie cellulaire de traduction : le ribosome. Ce dernier décode l'information génétique contenue dans les molécules d'ARNm qu'il convertit en protéines en utilisant des ARN de transfert (ARNt) aminoacylés. De nombreux ARNnc (sRNA, ARNtm) contribuent à réguler la traduction suivant des voies multiples et complexes (Duval *et al.*, 2015) qui peuvent aussi mobiliser des facteurs protéiques spécifiques ou le remodelage de la structure de l'ARNm en réponse à des signaux environnementaux (Picard *et al.*, 2012).

Il est impossible de faire ici, en quelques pages, une description exhaustive de la traduction bactérienne et de sa régulation. C'est pourquoi, je ne détaille ci-dessous que quelques aspects représentatifs dont, bien évidemment, ceux impliquant le facteur Rho ou d'autres ARN hélicases.

A. La synthèse des ARN impliqués dans la traduction

a. Les ARNr

Chez les bactéries, la grande sous-unité 50S du ribosome comprend les ARNr 5S et 23S tandis que la petite sous-unité 30S contient l'ARNr 16S (voir page 37). Les gènes codant pour ces ARNr sont regroupés sous la forme d'un opéron appelé *rrn*. La capacité de la bactérie à répondre aux variations de son environnement est en général corrélée avec le nombre de copies *rrn* dans son génome (Klappenbach *et al.*, 2000). Ceci lui permet d'ajuster rapidement la quantité d'ARNr synthétisés (la transcription des opérons *rrn* peut représenter ~ 70 % des évènements transcriptionnels chez *E. coli*) aux conditions et ressources disponibles et d'optimiser la vitesse de croissance (Stevenson & Schmidt, 1998).

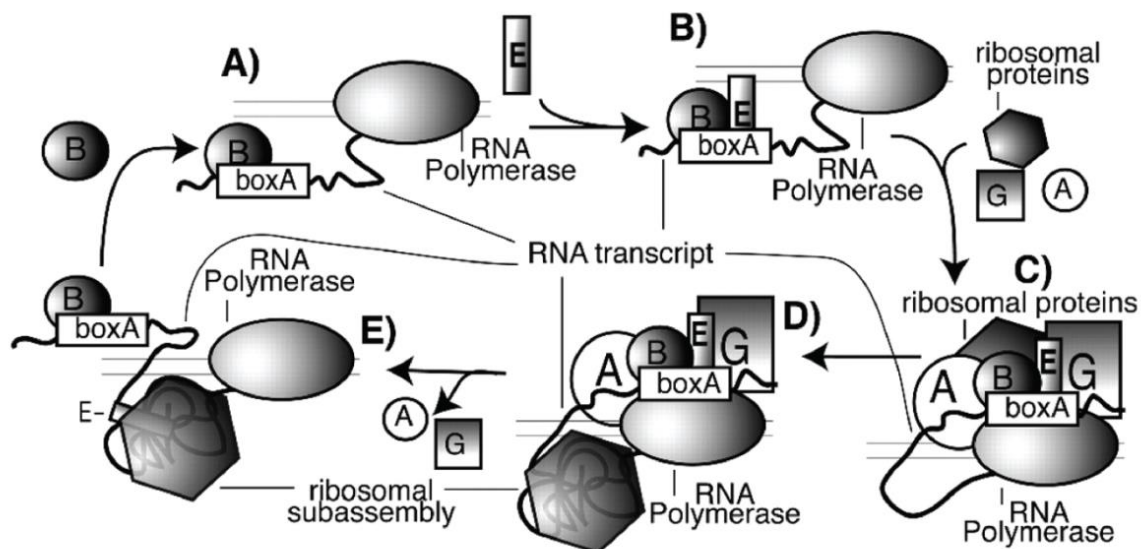


Figure 12 : Modèle d'assemblage du complexe d'anti-termination des opérons *rrn*. (A) La transcription par l'ARNP d'une séquence *boxA* provoque le recrutement et la fixation du facteur NusB. Cette interaction est stabilisée dans un second temps par le facteur NusE (alias S10) (B). Le complexe NusB-NusE-*boxA* induit le recrutement des autres protéines du complexe d'anti-termination (C) augmentant la stabilité du CET et régulant sa vitesse d'élongation pour provoquer le repliement correct des ARNr. Au cours de la transcription de l'ARNr, l'apparition des sites de fixation pour les protéines ribosomiques provoque leur réarrangement depuis le complexe d'anti-termination vers la sous-unité ribosomique (D). A la fin, le réarrangement de NusE participe au relargage de NusB de la séquence *boxA* pour son recyclage (E) (d'après Greive et al., 2005).

Les opérons *rrn* n'étant pas traduits, ils constituent *a priori* une cible idéale pour la terminaison Rho-dépendante (**Figure 1**) (Santangelo & Artsimovitch, 2011). Cette dernière est cependant limitée par la forte structuration des ARNr et l'absence de "bulles C>G" significatives (voir Chapitre III page 81) en leur sein (Alifano et al., 1991; Rabhi et al., 2010a). De plus, l'action de Rho est bridée par un mécanisme d'anti-termination qui contribue à garantir une production importante des 3 types d'ARNr. L'anti-termination repose sur la formation d'un complexe nucléoprotéique requérant la présence de motifs *nut-like* (encore appelés AT, pour "Anti-Termination", et composés des séquences *boxA* et *boxB*) dans le transcrit naissant et la participation des protéines NusA, NusB, NusE/S10, NusG, S1, S4, L3, L4 et L13 (Kaczanowska & Rydén-Aulin, 2007). NusB est le premier facteur recruté lors de la transcription de la séquence *boxA* située en amont (**Figure 12.A**). L'interaction NusB:*boxA* est ensuite stabilisée par l'intervention de NusE (**Figure 12.B**) (Greive et al., 2005). Ces étapes gouvernent le recrutement des autres facteurs afin de créer un complexe multi-protéique protégeant l'ARNP de la terminaison Rho-dépendante suivant un mécanisme encore mal défini (Santangelo & Artsimovitch, 2011). Ce complexe est néanmoins supposé améliorer la stabilité du CET et réguler la vitesse d'élongation de l'ARNP afin de favoriser la

configuration de l'ARNr requise pour la formation du ribosome (**Figure 12.D & E**) (Kaczanowska & Rydén-Aulin, 2007).

Ce système d'anti-termination semble être très conservé chez les bactéries. Ainsi, des séquences ressemblant aux motifs *nut-like* sont présentes dans les opérons *rrn* d'un grand nombre d'espèces différentes (Arnvig *et al.*, 2008). La complémentation des séquences *nut-like* d'*E.coli* par ces séquences phylodivergentes, permet l'anti-termination *in vivo* avec une efficacité similaire, à l'exception de celles provenant de *Thermotoga maritima* et de *M. tuberculosis* (Arnvig *et al.*, 2008). Un système similaire est employé par le phage λ pour inhiber la terminaison Rho-dépendante lors de la transcription de certains gènes viraux. Dans ce système c'est l'interaction entre la protéine virale λN et la séquence *nut* (pour "*N utilization*", d'où est dérivé le nom des motifs *nut-like*) contenue dans le transcrit viral naissant qui va initier le recrutement des facteurs Nus nécessaires à l'anti-termination (Santangelo & Artsimovitch, 2011).

b. Les ARNt

Les ARNt (sous forme aminoacylée) sont les "briques" essentielles utilisées par le ribosome pour former (allonger) la chaîne peptidique lors de la traduction des ARNm. Cependant, leur présence peut être requise pour d'autres fonctions, indépendantes de l'action du ribosome, telles que la synthèse du peptidoglycane (Filipe *et al.*, 2000) ou la résistance à certains antibiotiques (Klein *et al.*, 2009).

Dans les génomes bactériens, les gènes codant pour les ARNt sont retrouvés sous la forme de gènes isolés codant pour un seul ARNt, sous forme d'opérons codant pour plusieurs ARNt et finissant en général par un gène codant une protéine, et sous forme d'insertions dans les opérons *rrn* (Hori, 2001). Dans la majorité des cas, la synthèse des ARNt est achevée par l'action d'un terminateur intrinsèque se trouvant en aval du gène ou de l'opéron (Peters *et al.*, 2009; Shepherd & Ibba, 2015). Néanmoins, il fut montré très tôt que la transcription de l'opéron *tyrTV* d'*E. coli* est assujettie à la terminaison Rho-dépendante *in vitro* et *in vivo* (Küpfer *et al.*, 1978). Plus récemment, une cartographie des terminateurs Rho-dépendants putatifs d'*E. coli* a généralisé ce phénomène à 12 des 36 unités transcriptionnelles codant des ARNt (**Tableau 1**), en dehors de ceux contenus dans les opérons *rrn* (Peters *et al.*, 2009). La terminaison Rho-dépendante est proposée compenser l'inactivation de terminateurs intrinsèques due soit à une mutation, soit à l'insertion d'ADN étranger (Peters *et al.*, 2009).

Tableau 1 : Gènes et opérons codant des ARNt subissant une terminaison Rho-dépendante ^a

Opéron	Brin	Début	Fin	Termineur intrinsèque
<i>thrW</i>	+	262095	262170	-
<i>argU</i>	+	563946	564022	-
<i>tyrTV</i>	-	1286845	1286467	-
<i>valVW</i>	+	1744459	1744616	-
<i>serU</i>	-	2041581	2041492	-
<i>asnU</i>	+	2057875	2057950	+ (fonctionnel ?)
<i>argW</i>	+	2464331	2464405	-
<i>serV-argVYZQ</i>	-	2816667	2815806	-
<i>glyU</i>	-	2997079	2997006	-
<i>selC</i>	+	3834245	3834339	-
<i>glyVXY</i>	+	4390383	4390681	-
<i>leuQPV</i>	-	4604424	4604101	-

a : D'après (Küpper *et al.*, 1978; Peters *et al.*, 2009)

Les ARNt néo-synthétisés subissent une étape de maturation de leurs extrémités 5' et 3' qui va également permettre la séparation des ARNt provenant de transcrits polycistroniques (Gegenheimer & Apirion, 1981; Li & Deutscher, 1996; Ow & Kushner, 2002). Cette étape est assurée par différentes ribonucléases (RNases): RNase E et RNase III pour la maturation des extrémités 3' et le clivage des précurseurs polycistroniques et RNase P pour la maturation des extrémités 5' (Altman, 1971; Altman & Smith, 1971). Chez *E. coli*, la RNase P peut aussi prendre en charge le clivage de précurseurs polycistroniques et la maturation des extrémités 3' si la séquence du termineur intrinsèque a été préalablement éliminée (Agrawal *et al.*, 2014; Mohanty & Kushner, 2007; Mohanty & Kushner, 2008). Il est important de souligner que les ARNt dont la terminaison de la transcription est assurée par Rho sont maturés de manière différente, indépendamment de l'action de la RNase E (Mohanty & Kushner, 2008).

Dans une seconde phase de maturation, les ARNt subissent diverses modifications chimiques visant à augmenter l'efficacité de la traduction. Suivant les espèces, 1 à 10 % des gènes codent des protéines impliquées dans la modification des ARNt, soulignant l'importance de ce phénomène (El Yacoubi *et al.*, 2012; Motorin & Helm, 2010). L'ARNt est ensuite engagé dans une réaction d'aminacylation de son extrémité 3' (via le groupement 3'-OH ou 2'-OH) catalysée par une aminoacyl-tRNA synthétase spécifique. Une fois l'acide aminé chargé, l'ARNt sera pris en charge par le facteur d'élongation EF-Tu afin d'être acheminé au ribosome pour procéder à l'élongation de la chaîne peptidique. Certains ARNt

chargés peuvent subir des étapes alternatives de maturation ou d'édition que je n'aborderai pas ici (pour revue : Shepherd & Ibba, 2015).

B. Ribogénèse et composition des ribosomes

Les ribosomes sont de gros complexes RNP comprenant deux sous-unités : la grande sous-unité 50S contient les ARNr 5S et 23S ainsi que 33 protéines ribosomiques (L1-L7, L9-L11, L13-L25 et L27-L36) tandis que la petite sous-unité 30S comprend l'ARNr 16S et 21 protéines ribosomiques (S1-S21). Bien que les ribosomes puissent être assemblés *in vitro* sans apport de cofacteurs, ce processus nécessite des conditions très éloignées de celles du cytoplasme bactérien (Dohme & Nierhaus, 1976; Nierhaus, 1991; Traub & Nomura, 1968). *In vivo*, l'assemblage des ribosomes est vraisemblablement gouverné par l'ordonnancement co-transcriptionnel de la synthèse des ARNr (**Figure 12**) mais également par l'action de multiples cofacteurs (pour revue : Britton, 2009; Shajani *et al.*, 2011; Wilson & Nierhaus, 2007). Dans ce qui suit, je n'évoquerai que les protéines "DEAD-box", des (re)modeleurs de complexes RNP appartenant à une classe majeure d'ARN hélicases (voir Chapitre II page 51). A ce jour, quatre des cinq protéines "DEAD-box" d'*E. coli* (SrmB, DeaD, DbpA et RhlE) ont été impliquées dans la synthèse des ribosomes (Iost *et al.*, 2013; Redder *et al.*, 2015).

Le rôle exact de la première d'entre elles, SrmB, reste assez mal défini même s'il a été montré que la protéine est capable de se fixer sur un site proche de l'extrémité 5' de l'ARN 23S et de favoriser la formation de structures secondaires impliquées dans le recrutement des protéines L13 et L25 durant les étapes préliminaires d'assemblage du ribosome (Proux *et al.*, 2011). La délétion du gène *srmB* provoque chez *E. coli* l'accumulation de particules ribosomiques dénuées de plusieurs protéines et sédimentant autour de 40S (Charollais *et al.*, 2003).

La protéine DeaD (alias CsdA) est surexprimée en conditions de culture à basse température (Jones *et al.*, 1996). DeaD a été initialement caractérisée comme supprimeur de mutation de la protéine S2 et, donc, comme participant à la formation de la sous-unité 30S du ribosome (Toone *et al.*, 1991). Néanmoins, DeaD sédimente à faible température avec la sous-unité 50S et la délétion du gène *deaD* provoque la formation de particules 40S (Charollais *et al.*, 2004). DeaD semble jouer un rôle préférentiel dans la formation de la grande sous-unité mais un rôle, peut-être indirect, dans la formation de la petite sous-unité ne peut être exclu (Charollais *et al.*, 2004; Iost *et al.*, 2013; Kitahara & Suzuki, 2009). La

composition des particules 40S formées dans un mutant $\Delta deaD$ révèle un défaut en protéines ribosomiques intégrées dans les étapes tardives d'assemblage du ribosome (Peil *et al.*, 2008). Bien que les effets de la délétion $\Delta deaD$ soient observables à 37°C, ils sont plus marqués dans des conditions de croissance à faible température, suggérant que DeaD facilite la formation des ribosomes dans des conditions de froid.

DpbA (ou son homologue YxiN chez *B. subtilis*) peut se fixer, via son domaine CTD, à la tige-boucle 92 de l'ARNr 23S qui se trouve proche du site actif du ribosome (Fuller-Pace *et al.*, 1993; Kossen *et al.*, 2002). Néanmoins, le rôle de DpbA reste mal compris car la délétion du gène *dbpA* n'induit pas de phénotype particulier quelles que soient les conditions de culture (Elles & Uhlenbeck, 2008; Peil *et al.*, 2008; Sharpe Elles *et al.*, 2009). Une mutation inhibant l'activité enzymatique de DbpA mais pas sa capacité à fixer l'ARN provoque l'accumulation de particules 45S dépourvues de protéines proches du site actif du ribosome et contenant un ARNr 23S incorrectement replié (Elles & Uhlenbeck, 2008). Selon les auteurs, DpbA ne serait pas cruciale car sa fonction redondante pourrait être assurée par une voie alternative. La saturation du site de fixation par le mutants DpbA inhiberait par compétition la voie alternative, provoquant l'accumulation de particules incomplètes (Sharpe Elles *et al.*, 2009).

La protéine RhlE a un rôle ambigu dans l'assemblage des ribosomes. La surexpression de RhlE supprime ou aggrave, respectivement, le phénotype de sensibilité au froid des mutants $\Delta deaD$ et $\Delta srmB$ (Awano *et al.*, 2007; Jain, 2008). La délétion du gène *rhlE* a les effets exactement inverses sur les phénotypes des mutants $\Delta deaD$ et $\Delta srmB$ (Jain, 2008). Jain explique ces observations par un modèle où SmrB et DeaD sont impliquées dans deux voies parallèles d'assemblage du ribosome, différenciées par la conformation de l'ARN 23S (Jain, 2008). L'expression de RhlE favoriserait le repliement de l'ARN 23S engageant la voie SmrB, rendant le rôle de cette dernière plus crucial et celui de DeaD moins important. Néanmoins, ce modèle reste encore aujourd'hui débattu (Iost *et al.*, 2013).

C. Synthèse des protéines

Indépendamment du recyclage des ribosomes, la biosynthèse des protéines comprend trois phases principales: l'initiation, l'élongation et la terminaison (**Figure 13**). Chacune de ces phases est susceptible d'affecter le niveau d'expression génique, soit directement en modulant le nombre de ribosomes couvrant l'ARNm (qui constituent le polysome) (**Figure**

14.A & B) ou le nombre de protéines synthétisées par molécule d'ARNm, soit indirectement en impactant le couplage traduction/transcription (et le nombre de molécules d'ARNm synthétisées) ou l'accès des ARNm aux RNases (et le pool d'ARNm intègres disponibles ; (Deana & Belasco, 2005; Deneke *et al.*, 2013; Hussein *et al.*, 2015)).

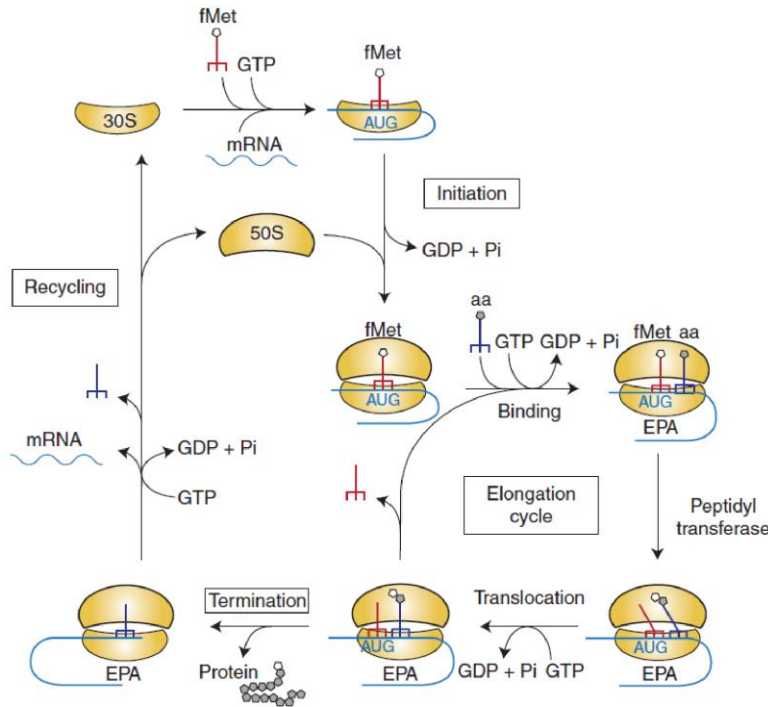


Figure 13 : Biosynthèse des protéines chez les bactéries.

Les cofacteurs permettant le déroulement de chaque étape ne sont pas représentés. A chaque cycle d'élongation, un nouvel acide aminé (pentagone gris) est ajouté à la chaîne peptidique. La terminaison intervient quand un codon stop est localisé dans le site A du ribosome, ce qui provoque l'hydrolyse de la chaîne peptidique dans le site P (d'après Hershey *et al.*, 2012).

L'initiation nécessite la fixation de la petite sous-unité ribosomique 30S sur la partie 5' non traduite (5'-leader) du transcrit ARN (**Figure 13**) (Laursen *et al.*, 2005). Cette fixation est favorisée par les facteurs d'initiation IF-1, IF-2 et IF-3 qui aident la sous-unité 30S à appairer la séquence SD avec la région anti-SD de l'ARN 16S. Cet appariement favorise le placement du codon d'initiation dans le site P du ribosome où il est alors reconnu par l'ARNt-fMet initiateur. La sous-unité 50S vient compléter ce complexe pré-initiateur (**Figure 13**) d'où sont relargués les facteurs IF1, IF2 et IF3 afin de former le ribosome 70S capable d'assurer l'élongation (Laursen *et al.*, 2005; Simonetti *et al.*, 2009). L'efficacité et la fréquence d'initiation sont fortement dépendantes de la concentration en ribosomes disponibles et de la séquence du *rbs* (comprenant entre autre le motif SD, le codon d'initiation et la région entre les deux) qui peut être plus ou moins optimale pour la fixation du ribosome, permettre des repliements compétitifs de l'ARNm et/ou contenir des sites d'interaction(s) avec des facteurs régulateurs (voir page 47) (Duval *et al.*, 2015; Hussein *et al.*, 2015; Laursen *et al.*, 2005).

Durant l'élongation, le codon ARNm situé dans le site A du ribosome va être apparié à son ARNt aminoacylé correspondant via l'intervention de la GTPase EF-Tu (Reynolds *et al.*, 2010) ; Reynolds *et al.*, 2010). Ceci engendre le transfert et l'addition de la fMet (ou de la chaîne peptidique naissante dans les étapes suivantes d'élongation) associée à l'ARNt présent dans le site P sur l'ARNt aminoacylé présent dans le site A (ce processus correspondant à une réaction de transamidification ester $\rightarrow$ amide ; **Figure 13**). La translocation GTP-dépendante du ribosome modifie alors le registre d'occupation des ARNt : du site A au site P pour l'ARNt portant le peptide naissant et du site P au site E pour l'ARNt "dégarni" d'où il sera relargué du ribosome (**Figure 13**). Le nouveau codon ARNm occupant le site A permettra d'amorcer un nouveau cycle d'élongation (Hershey *et al.*, 2012). Ce processus d'élongation de la chaîne peptidique intervient à une vitesse moyenne qui peut varier significativement (5-20 aa/s pour *E. coli* ;(Proshkin *et al.*, 2010; Spencer & Barral, 2012; Young & Bremer, 1976) en fonction de nombreux facteurs comme la séquence de l'ARNm (présence de codons rares, par exemple), le pool d'ARNt aminoacylés disponibles, voire la présence d'autres ribosomes sur le même ARNm (El Yacoubi *et al.*, 2012). Par exemple, la carence en acides aminés s'accompagne par une augmentation du taux d'ARNt non chargés dont la fixation EF-Tu indépendante au sein du ribosome provoque la mise en pause de celui-ci (Starosta *et al.*, 2014). Cet événement induit à son tour la mise en place de la réponse stringente via notamment le recrutement de l'enzyme RelA/SpoT par le ribosome en pause et la production de l'alarmone (p)ppGpp (Somerville & Ahmed, 1979; Wendrich *et al.*, 2002). Les effets de (p)ppGpp sont divers et provoquent un changement du métabolisme global de la bactérie afin de mieux s'adapter au stress (Zhou & Jin, 1998). On peut noter, néanmoins, que l'un des effets de l'alarmone est la diminution de la vitesse d'élongation (ou stimulation de la pause) de l'ARNP au niveau de certains gènes (voir page 92) (Dennis & Nomura, 1974; Washburn & Gottesman, 2015).

Lorsque la translocation du ribosome provoque l'entrée d'un codon stop (UAA, UAG ou UGA) dans le site A, un facteur de relargage (RF1 ou RF2) est recruté. Ce facteur est un mime d'ARNt qui provoque le clivage de l'ARNt portant la chaîne peptidique (par hydrolyse de la liaison ester) et le relargage de cette dernière dans le cytoplasme (Hershey *et al.*, 2012; Laursen *et al.*, 2005). Ce faisant, les sous-unités 30S et 50S se dissocient de l'ARNm et peuvent être recyclées pour une nouvelle synthèse protéique (Hershey *et al.*, 2012; Laursen *et al.*, 2005). L'efficacité de ce processus de terminaison est parfois modulée par des métabolites

qui affectent ainsi le temps de résidence du ribosome sur le codon stop, ce qui peut avoir également des conséquences pour la régulation transcriptionnelle (voir Chapitre III page 90).

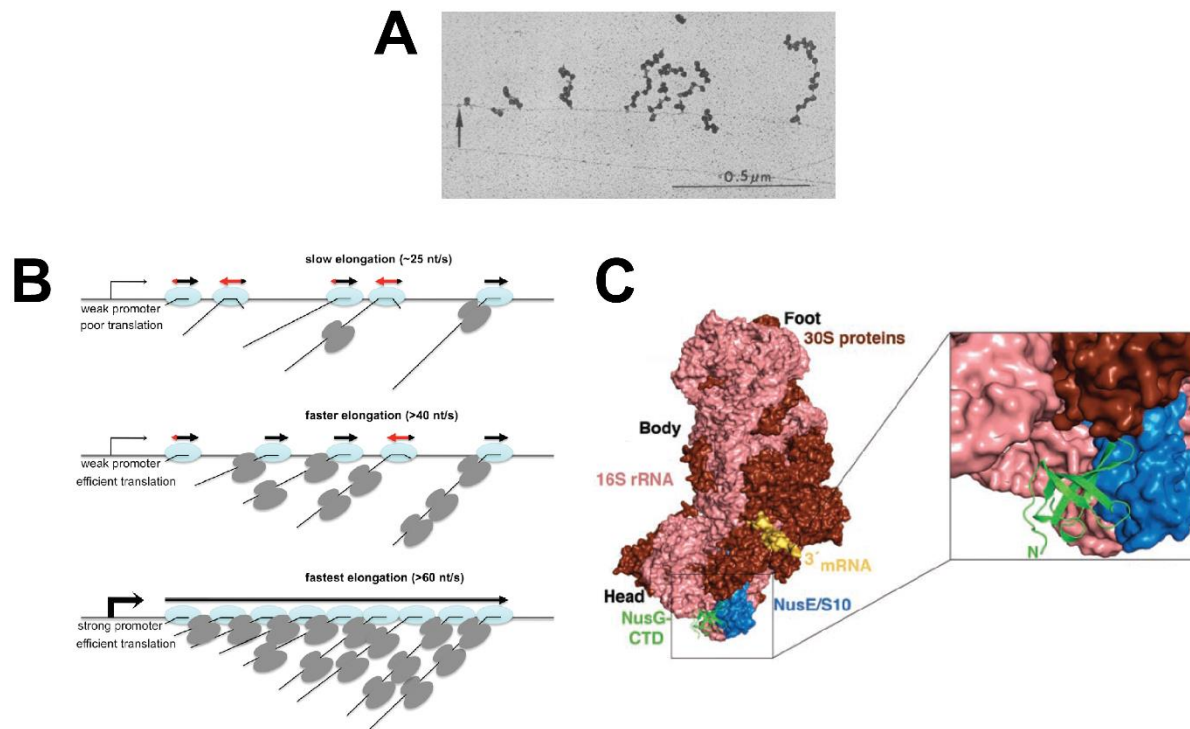


Figure 14 : La transcription et la traduction sont concomitantes et couplées chez les bactéries. (A) Photo de microscopie électronique de polysomes. La flèche indique le site putatif de l'initiation de la transcription (d'après (Miller et al., 1970)). (B) Modulation de la vitesse d'élongation de transcription par la "force" du couplage traduction/transcription (d'après (Proshkin et al., 2010)). (C) Représentation structurale de l'interaction entre le CTD de NusG (vert) et la protéine ribosomique NusE/S10 (bleu) (d'après Burmann et al., 2010).

L'efficacité globale de la traduction est liée (mais pas proportionnelle) à la taille du polyribosome (alias polysome) (Racle et al., 2013) dont le ribosome "leader" contacte directement l'ARNP (**Figure 14.B**). Ce couplage entre traduction et transcription concomitantes est unique aux bactéries (Miller et al., 1970). Il est "garanti" physiquement par le cofacteur NusG dont le NTD interagit avec l'ARNP et le CTD interagit avec la protéine NusE/S10 du ribosome "leader" (**Figure 14.C**) (Burmann et al., 2010) durant la transcription des séquences codantes. Ce couplage traduction/transcription, qui est une forme de coopération entre moteurs moléculaires, augmente la processivité de la transcription en *i*) limitant les pauses de l'ARNP et permettant à celle-ci de déplacer des obstacles placés sur l'ADN (Proshkin et al., 2010), *ii*) empêchant la terminaison Rho-dépendante (voir Chapitre III page 84) et *iii*) dans certains cas, en inhibant la terminaison intrinsèque (Abe et al., 1999; Wright & Hayward, 1987).

Si la transcription co-traductionnelle des ARNm apparaît être la norme, la traduction co-transcriptionnelle pourrait être beaucoup plus faible que la traduction d'ARNm matures advenant à distance du nucléoïde (Bakshi *et al.*, 2012; Bakshi *et al.*, 2015; Gottshall *et al.*, 2014). Cette forte traduction post-transcriptionnelle pourrait contribuer à expliquer pourquoi l'inhibition de la terminaison Rho-dépendante affecte beaucoup moins le protéome que le transcriptome d'*E. coli* (Cardinale *et al.*, 2008; Hussein *et al.*, 2015)

3. DÉGRADATION DES ARN

La dégradation des ARN est un composant essentiel du métabolisme cellulaire et prend une place importante dans le processus d'expression génique. Les mécanismes de dégradation sont diversifiés dans le règne bactérien et à l'origine d'une grande variabilité des durées de vie moyennes des ARN (Bandyra *et al.*, 2013; Lehnik-Habrink *et al.*, 2012). Cette diversification provient de la composition variable de la machinerie de dégradation principale – le dégradosome – mais aussi de la coexistence de différentes voies de dégradation dont la nature et/ou l'importance peuvent également varier significativement selon les espèces. Dans ce qui suit, je me contente de décrire (et différencier) brièvement les deux voies de dégradation principales retrouvées chez *E. coli* et *B. subtilis* (dénommées “voie 5' ” et “voie directe” par Laalami *et al.*, 2014) et n'évoque le cas d'autres espèces que lorsqu'un lien avec Rho a été établi.

A. La voie 5'-monophosphate

La “voie 5' ” est initiée par la conversion de l'extrémité 5'-triphosphate en extrémité 5'-monophosphate (Durand *et al.*, 2015; Laalami *et al.*, 2014). Chez *E. coli* et *B. subtilis*, cette conversion est assurée par la pyrophosphatase RppH (**Figure 15**) qui peut être inhibée par la présence de structures secondaires dans la région 5'-proximale de l'ARN (Celesnik *et al.*, 2007; Deana *et al.*, 2008; Hsieh *et al.*, 2013). Chez *E. coli* - et d'autres espèces comme *M. tuberculosis* ou *Aquifex aeolicus* (phylum *Aquificae*) - la formation d'une extrémité 5'-monophosphate stimule le recrutement de la RNase E (via son domaine senseur situé dans sa partie NTD catalytique), une endonucléase qui va cliver la chaîne ARN dans sa partie 5' (Bandyra *et al.*, 2013; Kaberdin & Bizebard, 2005; Zeller *et al.*, 2007). Ce clivage génère un nouveau substrat ARN 5'-monophosphate pour la RNase E qui, par son action répétée, va tronçonner l'ARN en de multiples fragments (**Figure 15.A**, panneau de droite) (Laalami *et*

al., 2014). Ces fragments seront pris en charge par une 3'-exonucléase, la PNPase, pour être complètement dégradés (**Figure 15.A**). Certains fragments peuvent subir une polyadénylation facilitant leur dégradation par la PNPase, notamment lorsqu'ils possèdent une extrémité 3' fortement structurée (Régnier & Hajnsdorf, 2009). Chez *B. subtilis*, la conversion de l'extrémité 5'-triphosphate en extrémité 5'-monophosphate stimule le recrutement de l'hétérocomplexe RNase J1/J2 qui va directement dégrader l'ARN en nucléotides via son activité 5'-exonucléase (**Figure 15.B**, panneau de droite) (Durand *et al.*, 2015; Laalami *et al.*, 2014).

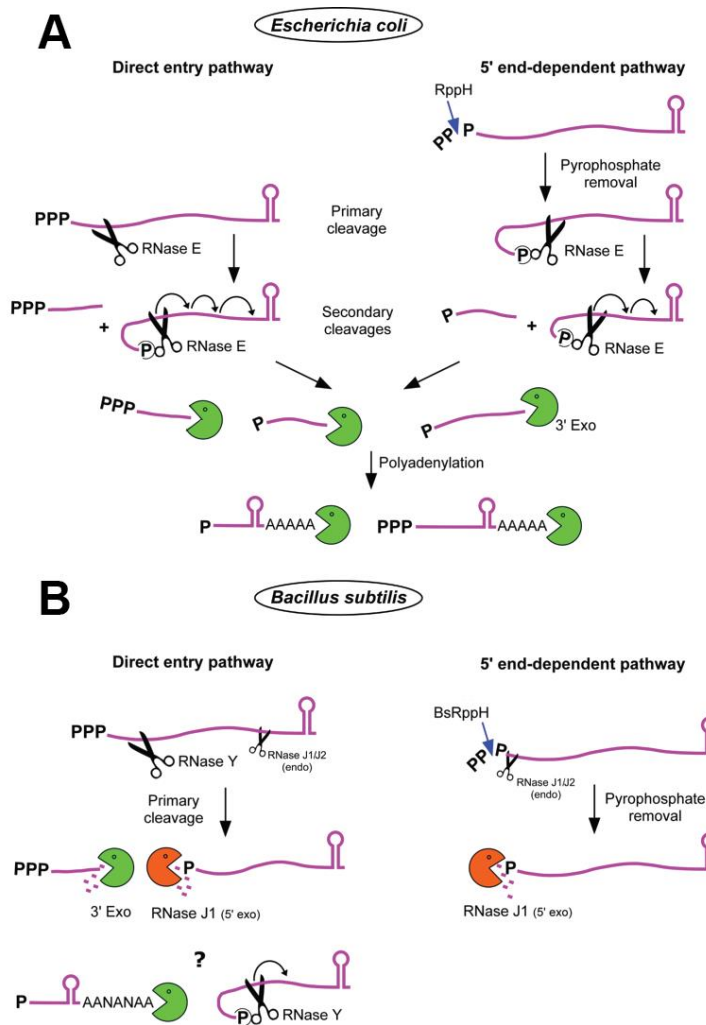


Figure 15 : Comparaison des voies de dégradation de l'ARN chez *B. subtilis* et *E. coli*.

Chez *E. coli*, les deux voies principales de dégradation impliquent de multiples coupures par la Rnase E permettant la dégradation par l'activité 3' exonucléase de la PNPase (vert). Chez *B. subtilis*, la dégradation par l'exonucléase J1 nécessite soit le clivage préalable de l'ARN (voie directe), soit la conversion de l'extrémité 5'-triphosphate en 5'-monophosphate (voie 5') (d'après Laalami *et al.*, 2014).

B. La voie directe

La dégradation de certains transcrits est peu affectée par le statut de l'extrémité 5' (Bandyra *et al.*, 2013). Ces transcrits sont généralement dégradés par une voie alternative –la “voie directe”– qui, aussi bien chez *E. coli* que chez *B. subtilis*, implique le clivage préalable du transcrit par une endonucléase. Il est important dans ce cas que le transcrit ne soit pas

traduit car la présence de ribosomes pourrait alors masquer le site initial de coupure (Dreyfus, 2009). Chez *E. coli*, l'endonucléase concernée est la RNase E (**Figure 15.A**, panneau de gauche) qui peut utiliser son CTD pour interagir avec les transcrits au niveau de séquences riches en A/U (Jiang *et al.*, 2000). Pour certains transcrits, le clivage initial par la RNase E s'inscrit dans un processus de maturation (Bandyra *et al.*, 2013). Dans la plupart des cas, néanmoins, les fragments ARN résultant de ce clivage initial feront l'objet de coupures subséquentes par la RNase E puis de dégradation par la PNPase suivant les mêmes principes que pour la voie 5' (**Figure 15.A**) (Laalami *et al.*, 2014). Chez *B. subtilis*, le clivage initial du transcrit peut mobiliser la RNase Y au niveau d'une séquence simple-brin riche en A/U, la RNase III au niveau d'une structure tige-boucle, voire l'activité endonucléase du complexe RNase J1/J2 (**Figure 15.B**, panneau de gauche). Les fragments générés seront ensuite dégradés en nucléotides par les activités exonucléase de la PNPase et du complexe RNase J1/J2 (**Figure 15.B**) (Durand *et al.*, 2015).

C. Implication des hélicases dans le dégradosome

Il est maintenant admis que la dégradation des ARN n'est pas le fruit de l'action d'enzymes isolées (comme pourrait le laisser penser le paragraphe précédent) mais que celles-ci opèrent au sein de gros complexes multi-protéiques dont le plus important est appelé "dégradosome" (**Figure 16.A**).

Le composant principal du dégradosome d'*E. coli* et de nombreuses autres espèces bactériennes est la RNase E qui comporte des domaines accessoires (et phylogénétiquement peu conservés) assurant son ancrage à la membrane bactérienne et lui permettant de servir de plateforme d'interaction pour de nombreux cofacteurs (**Figure 16.A**) (Marcaida *et al.*, 2006). Chez *B. subtilis* et les espèces apparentées dépourvues de RNase E, c'est la RNase Y qui est fonctionnellement la plus proche (**Figure 15**) et pourrait jouer ce rôle de plateforme (**Figure 16.A**) (Lehnik-Habrink *et al.*, 2012). Ce point de vue reste néanmoins débattu (Bandyra *et al.*, 2013).

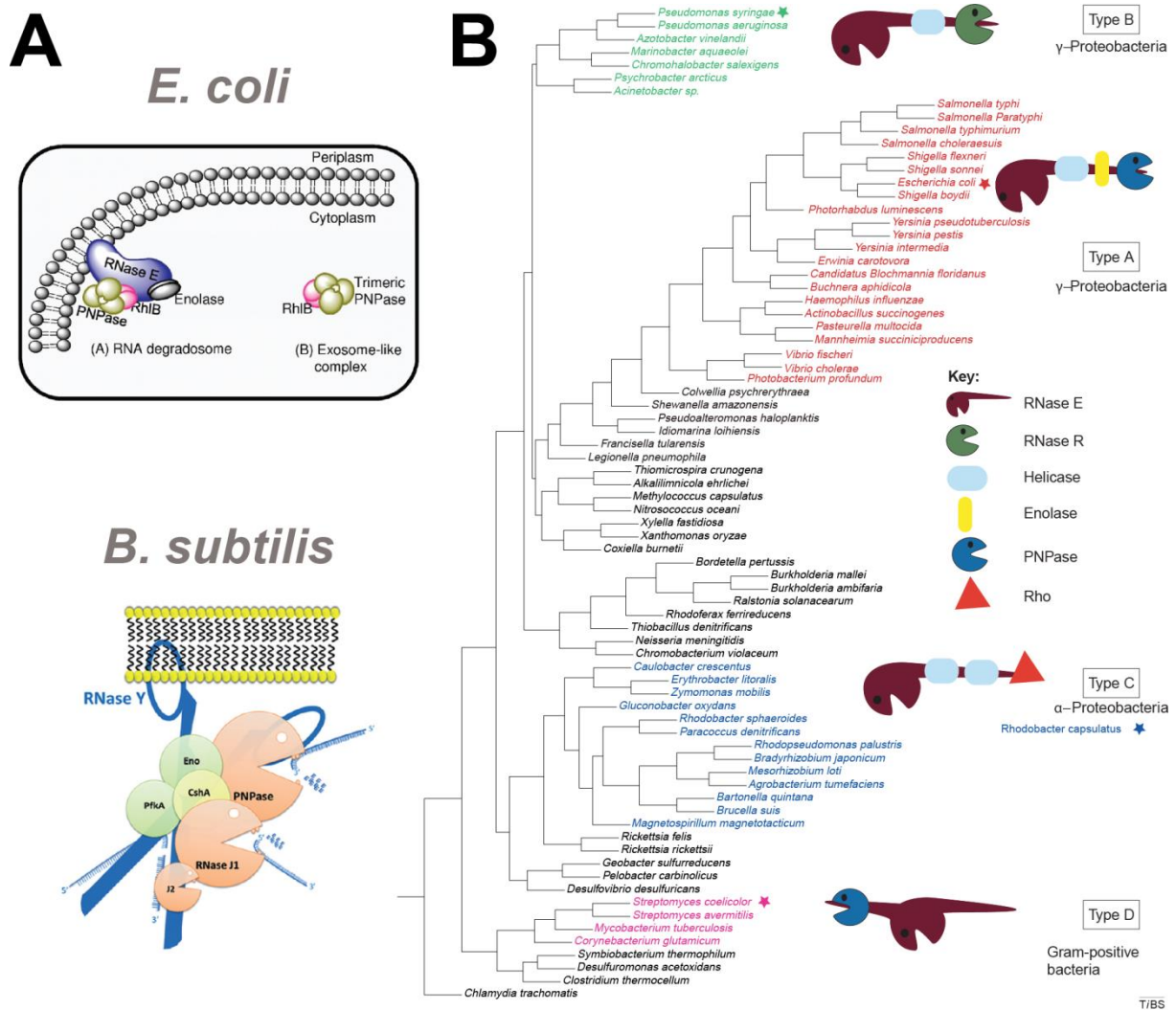


Figure 16 : Le dégradosome RNase E-dépendant au sein de la diversité bactérienne. (A) Représentation schématique des principaux complexes protéiques impliqués dans la dégradation et le processing des ARN chez *E. coli* et *B. subtilis* (chez qui la RNase E est probablement remplacée par la RNase Y). Le dégradosome est ancré à la membrane (d'après Kaberdin et al., 2011; Lehnik-Habrink et al., 2012). (B) La composition du dégradosome RNase E-dépendant varie significativement suivant les espèces. Les noms en rouge, vert, bleu et rose correspondent aux espèces dont le dégradosome est supposé se conformer aux types A, B, C et D schématisés sur la figure. Les espèces qui, comme *B. subtilis*, sont dépourvues de RNase E ne sont pas représentées (d'après Marcaida et al., 2006).

Quelle que soit l'espèce bactérienne, il semble que le dégradosome contient toujours au moins une hélicase DEAD-box (**Figure 16**) dont le rôle serait de faciliter la dégradation des ARN structurés, notamment au niveau de l'extrémité 3' (Chandran et al., 2007; Lehnik-Habrink et al., 2012). Chez *E. coli*, cette hélicase est RhIB dans les conditions normales de pousse (Khemici et al., 2008; Py et al., 1996; Redder et al., 2015) bien que la délétion de *rhlB* n'ait que peu d'effet sur le transcriptome (Bernstein et al., 2004). Il est possible que l'absence de RhIB puisse être compensée par une autre protéine DEAD-box ayant une fonction/activité redondante. Ainsi, en conditions de pousse à basse température, DeaD copurifie avec la RNase E et semble pouvoir suppléer RhIB dans le dégradosome (Prud'homme-Généreux et

al., 2004). Cette observation suggère que, pour la même espèce, la composition du dégradosome peut varier afin de mieux répondre aux conditions environnementales. Dans cette optique, il est intéressant de noter que Rho n'est généralement pas associé au dégradosome d'*E. coli* (**Figure 16.B**) (Marcaida *et al.*, 2006) mais co-purifie avec la RNase E et la PNPase au sein d'un complexe protéique dont les composants recherchés étaient DosC et DosP, des senseurs de l'oxygène (Tuckerman *et al.*, 2011). De façon intrigante, Rho semble être toujours associé au dégradosome chez *Rhodobacter capsulatus* (**Figure 16.B**) (Jäger *et al.*, 2001) bien qu'il soit retrouvé en plus grandes proportions dans des conditions de pousse en aérobie (Jäger *et al.*, 2004).

Un élément important contribuant à la stabilité des ARN et à leur prise en charge par la machinerie de dégradation est la polyadénylation de leur extrémité 3' (Régnier & Hajnsdorf, 2013). La plupart des ARNm (mais également certains ARNnc) d'*E. coli* sont polyadénylés dans des proportions (0,011-40%) et des longueurs (1-50 nt) variables (Mohanty *et al.*, 2004; Mohanty & Kushner, 2006) qui vont moduler la sensibilité des extrémités 3' à l'action de la PNPase et de deux autres 3'-exonucléases, la RNase II et la RNase R (Régnier & Hajnsdorf, 2013). Un élément central de la compétition dynamique entre allongement (par polyadénylation) et raccourcissement (par activités 3'-exonucléase) est la présence d'une structure secondaire stabilisatrice à l'extrémité 3' de l'ARN. Ainsi, les transcrits se terminant par un motif de terminaison intrinsèque sont des substrats préférentiels pour la poly(A) polymérase I chez *E. coli* (Régnier & Hajnsdorf, 2013). Dans ce contexte, on pourrait craindre que les transcrits issus de terminaison Rho-dépendante (ou, du moins, leurs extrémités 3') soient particulièrement instables. Néanmoins, les transcrits dénués de terminateur intrinsèque semblent également faire l'objet d'une extension de leur extrémité 3' reposant sur l'activité (secondaire) de polymérisation de la PNPase (Mohanty & Kushner, 2006). Dans ce cas, un long hétéropolymère (30-600nt) est ajouté à extrémité 3' par la PNPase (Mohanty *et al.*, 2004; Mohanty & Kushner, 2006). Ainsi, il est clair que de nombreux acteurs, (certains dont le rôle auxiliaire – Hfq, par exemple – ne peut être détaillé ici, faute de place) contribuent à la dégradation des ARN de manière contextuelle et multifactorielle. Globalement, la façon dont ce processus est coordonné avec les autres étapes du métabolisme ARN reste assez mystérieuse et susceptible de varier significativement selon le type d'ARN et selon les espèces.

4. LA REGULATION DE L'EXPRESSION GENIQUE DANS LA REGION 5'-LEADER DES ARN_M

Dans ce paragraphe, j'évoque brièvement le cas spécifique de la région 5' non-traduite (alias 5'-leader) des ARNm qui est une cible préférentielle pour la régulation, avec des effets documentés sur la transcription, la traduction et/ou la dégradation du transcrit associé (Marzi *et al.*, 2008; Oliva *et al.*, 2015; Van Assche *et al.*, 2015).

Tout d'abord, les régions 5'-leader sont des lieux de fixation de nombreuses protéines RBP (*RNA Binding Protein*), généralement au niveau de séquences ARN bien spécifiques (pour revue : Van Assche *et al.*, 2015). Dans certains cas, la fixation de la RBP provoque un remodelage structural de la région 5'-leader qui peut masquer (ou découvrir) un terminateur transcriptionnel ou un site *rbs*. Par exemple, la protéine TRAP de *B. subtilis* ou les protéines à domaine ANTAR retrouvées chez diverses espèces modulent la formation de terminateurs intrinsèques (Shu & Zhulin, 2002) alors que la protéine DEAD-Box d'*E. coli* DeaD semble nécessaire pour fondre des structures secondaires séquestrant le SD de certains gènes à basse température (Redder *et al.*, 2015). Un cas différent concerne la protéine CsrA (ou RsmA) qui est largement présente dans la diversité bactérienne (et pour lequel on trouve des homologues chez certains phages et eucaryotes ; <http://pfam.xfam.org>) et qui constitue un régulateur post-transcriptionnel majeur (>20% des ARNm d'*E. coli* interagissent avec CsrA) (Holmqvist & Vogel, 2013; Romeo *et al.*, 2013). Parce qu'elle reconnaît préférentiellement les séquences contenant un motif GGA (Dubey *et al.*, 2005), CsrA est supposée se fixer largement au niveau des séquences SD où elle serait un compétiteur pour les ribosomes (Baker *et al.*, 2002; Mercante *et al.*, 2009). Dans certains cas, CsrA pourrait masquer le codon AUG (Jonas *et al.*, 2008) ou s'associer à d'autres sites, plus en amont du 5'-leader (Yakhnin *et al.*, 2011). Parfois, la fixation de CsrA peut avoir un effet positif sur la traduction en empêchant la dégradation du transcrit (Yakhnin *et al.*, 2013). Il est intéressant de noter que CsrA peut aussi réguler l'expression d'un gène au niveau transcriptionnel. Chez *Pseudomonas aeruginosa*, par exemple, RsmA (l'orthologue de CsrA) est capable de diminuer la transcription du gène *lipA* suivant un mécanisme qui n'est pas encore été élucidé (Zha *et al.*, 2014). Chez *E. coli*, CsrA réprime l'expression de l'opéron *pgaABCD* en se fixant à la région 5'-leader de l'ARNm (voir Article III, page 151) mais aussi en réprimant *nhaR* qui code pour un facteur d'initiation de la transcription activant *pgaABCD* (Pannuri *et al.*, 2012). Enfin, le facteur Rho peut aussi être considéré comme une RBP capable de se lier aux régions 5'-leader des ARNm. Environ 1%

du transcriptome Rho-dépendant d'*E. coli* semble correspondre à des événements de terminaison localisés dans les régions 5'-UTR (**Figure 11**) (Peters *et al.*, 2012). Les régions 5'-leader des ARNm sont aussi des lieux de fixation de nombreux sRNA qui le plus souvent vont masquer les sites de reconnaissance/fixation du ribosome, inhiber la traduction et/ou favoriser la dégradation des ARNm cibles (Duval *et al.*, 2015; Geissmann *et al.*, 2009; Marzi *et al.*, 2008). Un phénomène de “polarité transcriptionnelle” Rho-dépendante est également une conséquence possible, comme cela a été observé pour l'opéron ChiPQ de *Salmonella enterica* (voir Chapitre III, pages 85 et 91 pour plus de détails) (Bossi *et al.*, 2012). Chez de nombreuses espèces, l'appariement entre le sRNA et son ARNm cible n'est pas suffisamment stable sans la participation de la protéine Hfq (Jousselin *et al.*, 2009), une ARN chaperonne hexamérique possédant de multiple sites de liaison à l'ARN (Figuerola-Bossi *et al.*, 2009; Geissmann & Touati, 2004; Vogel & Luisi, 2011). Les mécanismes moléculaires impliquant Hfq sont multiples et complexes et ne peuvent être détaillés ici, faute de place. On peut noter, cependant, que Hfq peut parfois directement se fixer à l'ARNm (en absence de sRNA) et promouvoir sa dégradation (Régner & Hajnsdorf, 2013; Vogel & Luisi, 2011) ou bien, encore, inhiber la terminaison Rho-dépendante (voir Chapitre III page 91) (Rabhi *et al.*, 2011a).

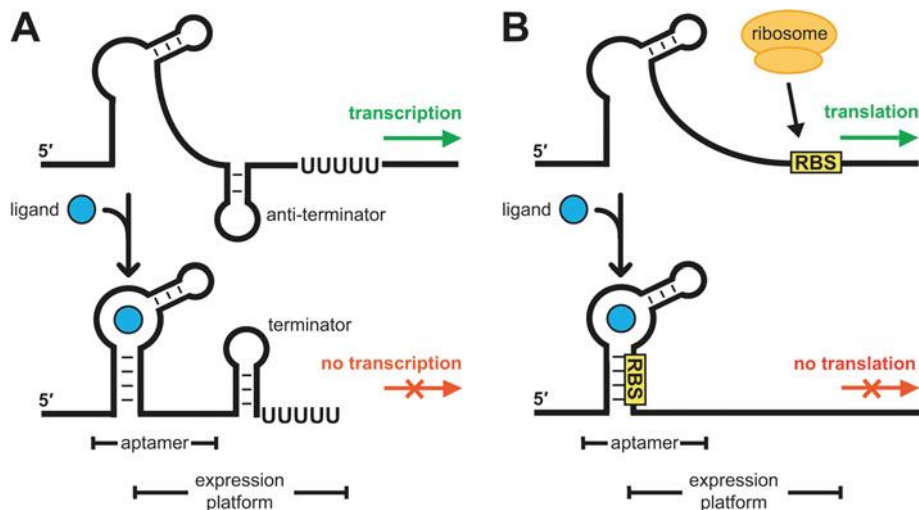


Figure 17 : Principe général de fonctionnement des motifs “riboswitch” (d’après Kim & Breaker, 2008).

Un dernier mécanisme de régulation significatif impliquant les régions 5'-leader des ARNm mobilise des motifs ARN particuliers appelés “riboswitch” (Mandal & Breaker, 2004). Ces motifs sont composés d'un motif “aptamère” qui, au travers la reconnaissance et la fixation d'un ligand particulier (en général un métabolite fonctionnellement lié à l'opéron/gène régulé), va moduler la conformation d'un second motif structural servant de “plateforme” d'expression (**Figure 17**). Suivant le principe déjà énoncé pour les RBP, le

remodelage (RBP- ou aptamère-dépendant) de la plateforme d'expression va séquestrer (ou libérer) un terminateur intrinsèque ou un site *rbs* (**Figure 17**), ce qui module l'expression des gènes situés en aval (Caldelari *et al.*, 2013; Duval *et al.*, 2015). Une analyse de la distribution des différents motifs “riboswitch” au sein de la diversité bactérienne a suggéré que ces mécanismes sont plus importants chez les Firmicutes où ils agiraient préférentiellement en atténuant la transcription (Barrick & Breaker, 2007). Depuis, il a été découvert chez *E. coli* et *S. enterica* (γ -Protéobactéries) ainsi que chez *Corynebacterium glutamicum* (Actinobactéries), des motifs “riboswitch” dont la plateforme d'expression gouverne un terminateur Rho-dépendant (Hollands *et al.*, 2012; Takemoto *et al.*, 2015). Il a aussi été trouvé des motifs riboswitch gouvernant la formation d'un ribozyme (Winkler *et al.*, 2004). Parfois, la conformation de la plateforme d'expression n'est pas modulée par la fixation d'un ligand (ou d'une RBP) mais par un stimulus physicochimique comme la température (Giuliodori *et al.*, 2010) ou le pH (Nechooshtan *et al.*, 2009).

Comme nous l'avons vu dans ce premier chapitre, la régulation fine du métabolisme ARN (et de l'expression génique), essentielle à la croissance et la survie de la bactérie, repose sur des mécanismes complexes et interdépendants mobilisant de nombreux cofacteurs. Parmi ces cofacteurs, les hélicases, dont notamment le facteur Rho, semblent jouer un rôle important. Ce groupe d'enzymes, aux structures et mécanismes moléculaires variés, est décrit brièvement dans le chapitre suivant.

Chapitre II : Les ARN hélicases

Les hélicases sont des protéines capables de dissocier des duplexes d'acides nucléiques ou de remodeler des complexes nucléoprotéiques en utilisant un mécanisme NTP-dépendant. On estime aujourd'hui que 1 % des génomes séquencés codent pour des hélicases et que ces enzymes participent à toutes les étapes du métabolisme impliquant des acides nucléiques (Byrd & Raney, 2012; Jankowsky & Fairman-Williams, 2010; Jarmoskaite & Russell, 2014; Patel & Picha, 2000; Redder *et al.*, 2015). Chez l'homme, un défaut de fonction (ou d'abondance) de certaines hélicases peut contribuer au développement de pathologies graves comme des cancers (Brosh, 2013; Robert & Pelletier, 2013; Steimer & Klostermeier, 2012).

Le terme "hélicase" regroupe un grand nombre d'enzymes aux fonctions et structures différentes qui sont classées selon divers critères. D'un point de vue structural, on distingue deux grands groupes d'hélicases qui peuvent être subdivisés en superfamilles (SF), elles-mêmes constituées de plusieurs familles. Le premier groupe comprend les hélicases de type monomérique, et est constitué des superfamilles SF1 et SF2 (**Tableau 2**). Le second groupe est constitué d'hélicases de type hexamérique et regroupe les superfamilles SF3, SF4, SF5, SF6 et SF7 (**Tableau 2**). Bien que certains spécimens aient une activité enzymatique quel que soit l'acide nucléique (ADN ou ARN) utilisé comme substrat, on distingue généralement les ADN hélicases des ARN hélicases qui sont le sujet de ce chapitre.

Tableau 2 : Classification des ADN et ARN hélicases

Superfamille	ARN hélicase	Etat oligomérique	Motif	Directionnalité
SF 1	Famille Upf1-Like	Monomérique ou dimérique	RecA	5'-3'
	Famille DEAD-box			Aucune
	Famille DEAH / RHA			5'-3' ou 3'-5'
SF 2	Famille RIG-I-Like	Monomérique ou dimérique	RecA	3'-5'
	Famille Ski2-Like			3'-5'
	Famille NS3 / NPH-II			3'-5'
SF 3	LT-Ag SV40 ?	Hexamérique	AAA+	3'-5'
SF 4	P4	Hexamérique	RecA	5'-3'
SF 5	Facteur Rho	Hexamérique	RecA	5'-3'
SF 6		Hexamérique	AAA+	3'-5'
SF 7		Hexamérique	AAA+	5'-3'

1. LES HÉLICASES MONOMÉRIQUES

Les hélicases faisant partie des SF1 et SF2 sont dites monomériques car les études biochimiques et biophysiques ont démontré que l'activité de translocation et/ou d'ouverture de régions en double brins (ci-dessous nommée “séparation de brins”) pouvait être supportée par un seul protomère (Dumont *et al.*, 2006; Galletto & Tomko, 2013; Hopfner & Michaelis, 2007). On note cependant que la formation de dimères, ou la coopération entre plusieurs monomères, peut parfois augmenter leur activité (Byrd & Raney, 2004; Galletto & Tomko, 2013; Serebrov *et al.*, 2009; Yokota *et al.*, 2013). Les ARN hélicases monomériques sont retrouvées dans six familles distinctes (**Figure 18 & Tableau 2**) sur la base d'homologies de leurs séquences (Fairman-Williams *et al.*, 2010).



Figure 18 : Répartition des ARN hélicases “monomériques” (indiquées par un R rouge) au sein des SF1 et SF2 (d’après E. Jankowsky, www.rnahelicase.org)

A. Organisation structurale des ARN hélicases “monomériques”

Il existe un grand nombre d'ARN hélicases monomériques et donc une grande diversité de séquences. Néanmoins, les études menées à ce jour démontrent que les organisations structurales de ces enzymes sont très similaires surtout au niveau du “cœur” hélicase, siège de l'activité de séparation de brins. Ce “cœur” hélicase est composé de deux domaines (D1 et D2), reliés par un linker, qui adoptent des conformations de type RecA (Berger, 2008; Cordin *et al.*, 2006) et forment une poche permettant la fixation et l'hydrolyse d'un NTP (**Figure 19 & Figure 20**) (Pyle, 2008; Singleton *et al.*, 2007). A l'intérieur des domaines D1 et D2, on retrouve plusieurs motifs conservés dont le nombre, initialement de sept (Gorbalenya & Koonin, 1993), n'a cessé d'augmenter (14 actuellement) avec la découverte de nouvelles ARN hélicases et le raffinement de leur classification phylogénétique

(Fairman-Williams *et al.*, 2010). Les différentes structures disponibles montrent que les motifs Q, I, II, et VI sont responsables de la fixation et de l'hydrolyse du NTP (**Figure 19** ;(Bono *et al.*, 2006; Caruthers *et al.*, 2000; Cordin *et al.*, 2004; Kim *et al.*, 1998; Sengoku *et al.*, 2006; Tortorici *et al.*, 2015; von Moeller *et al.*, 2009). Les structures d'hélicases SF1 et SF2 co-cristallisées avec un acide nucléique indiquent que la fixation du brin ARN se fait via des résidus présents dans les motifs Ia, Ib, IV et V et situés dans la partie supérieure des domaines D1 et D2 (**Figure 19**) (Caruthers *et al.*, 2000; Kim *et al.*, 1998; Sengoku *et al.*, 2006; von Moeller *et al.*, 2009). Enfin des études biochimiques ont montré que certaines mutations de résidus clés des motifs III et Va n'affectent pas ou peu la capacité des enzymes à hydrolyser les NTP ou à se fixer sur un brin ARN mais affectent l'activité hélicase (Caruthers *et al.*, 2000; Karow & Klostermeier, 2009; Rocak *et al.*, 2005). Ces résultats suggèrent une implication de ces motifs dans le couplage entre l'hydrolyse de l'ATP et l'activité de séparation des brins (Banroques *et al.*, 2008; Banroques *et al.*, 2010).

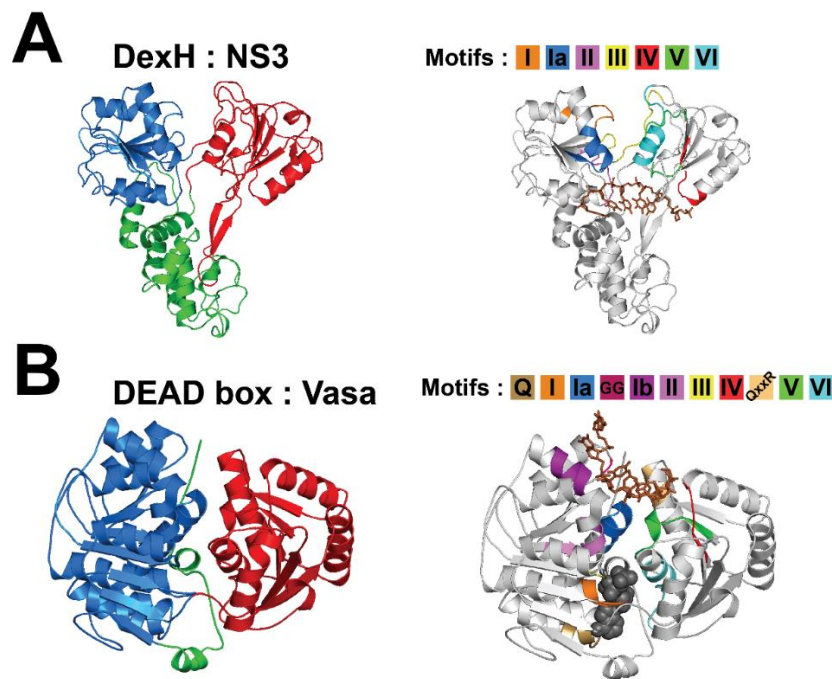


Figure 19 : Les différentes organisations structurales des hélicases monomériques. Structures de deux ARN hélicases appartenant à la SF2 (**A**) NS3 (PDB : 1A1V) et (**B**) Vasa (PDB : 2DB3). Le panneau de gauche montre les domaines de type RecA, D1 et D2 en bleu et rouge respectivement ainsi que le domaine supplémentaire CTD de la protéine en vert. Le panneau de droite montre les différents motifs impliqués dans la fixation de l'ARN et/ou l'interaction avec l'ATP qui sont représentés de différentes couleurs. Les oligonucléotides ARN co-cristallisés sont représentés en marron. L'analogue AMP-PNP co-cristallisé avec Vasa est représenté en boules noires.

En plus de ce “cœur” catalytique relativement bien conservé, les hélicases SF1 et SF2 possèdent souvent des domaines spécifiques (**Figure 20**). Bien que superflus pour l'activité hélicase *per se*, ces domaines peuvent être requis pour réguler la fonction de ces enzymes via

des interactions avec des partenaires protéiques (Chakrabarti *et al.*, 2011; Fiorini *et al.*, 2013; Kuang *et al.*, 2004; Rozen *et al.*, 1990), la reconnaissance de séquences ou structures particulières dans le substrat ARN (voir page 38) (Fuller-Pace *et al.*, 1993; Karow & Klostermeier, 2010; Kossen *et al.*, 2002; Rudolph & Klostermeier, 2009), l'orientation de la translocation (Myong *et al.*, 2007), ou une éventuelle oligomérisation (Klostermeier & Rudolph, 2009).

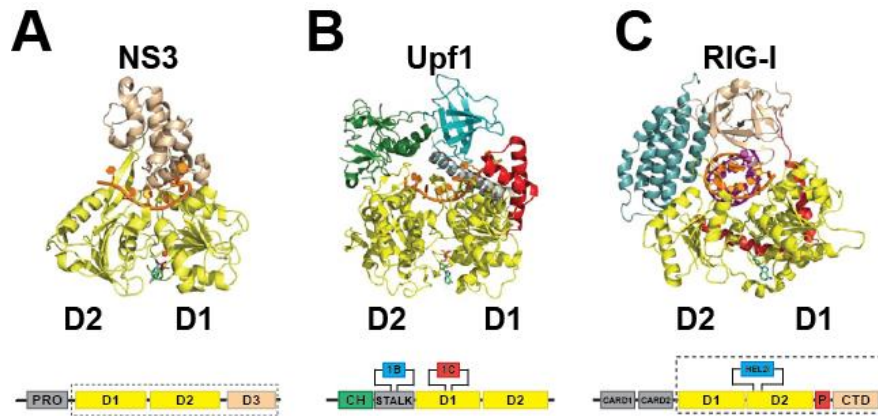


Figure 20 : Localisation des domaines accessoires chez différents spécimens SF1/SF2. Dans les panneaux A et C, Seules les régions encadrées en pointillé gris sont représentées dans la structure pour soucis de clarté. Les domaines D1 et D2 composant le “cœur” hélicase sont représentés en jaune et les domaines supplémentaires sont coloré en rose, bleu et/ou rouge (adapté de Ding & Pyle, 2012).

B. Mécanismes

Malgré leurs similarités structurales, les hélicases SF1 et SF2 ont des modes de fonctionnement qui peuvent être significativement différents. Certaines d'entre elles sont des translocases possédant une activité de séparation de brins dépendante de l'hydrolyse de NTP. Cette translocation est unidirectionnelle, $5' \rightarrow 3'$ ou $3' \rightarrow 5'$ selon les enzymes (Fairman-Williams *et al.*, 2010), et découle de la succession de réarrangements entre les domaines D1 et D2 induits par la fixation, l'hydrolyse et le relargage du nucléotide cofacteur (**Figure 21.A**) (Myong *et al.*, 2007). Un prototype de translocase SF2 est l'ARN hélicase NS3 du virus HCV. De nombreuses études biochimiques (Levin *et al.*, 2005), structurales (Appleby *et al.*, 2011; Gu & Rice, 2010) et biophysiques à l'échelle de la molécule unique (Dumont *et al.*, 2006; Myong *et al.*, 2007) supportent un mécanisme de translocation $3' \rightarrow 5'$ par pas de 1 nt/ATP hydrolysé de type “Brownian ratchet” (comme pour le CET ; voir **Figure 5**) mais au cours duquel une partie des contacts entre NS3 et l'ARN ne sont remodelés que tous les ~ 3 nt. Cela se traduit par une accumulation d'énergie dissipée périodiquement (tous les ~ 3 nt) (“spring-loaded mechanism” - **Figure 21.B**) (Ding & Pyle, 2012). On peut noter

également que l'état de plus forte (vs faible) liaison de NS3 à l'ARN est celui dans lequel la poche ATPase est libre (vs occupée par l'ATP). L'activité de séparation de brins de NS3 résulte d'un phénomène d'exclusion stérique, le site catalytique ne pouvant accommoder qu'un simple brin (Appleby *et al.*, 2011; Gu & Rice, 2010).

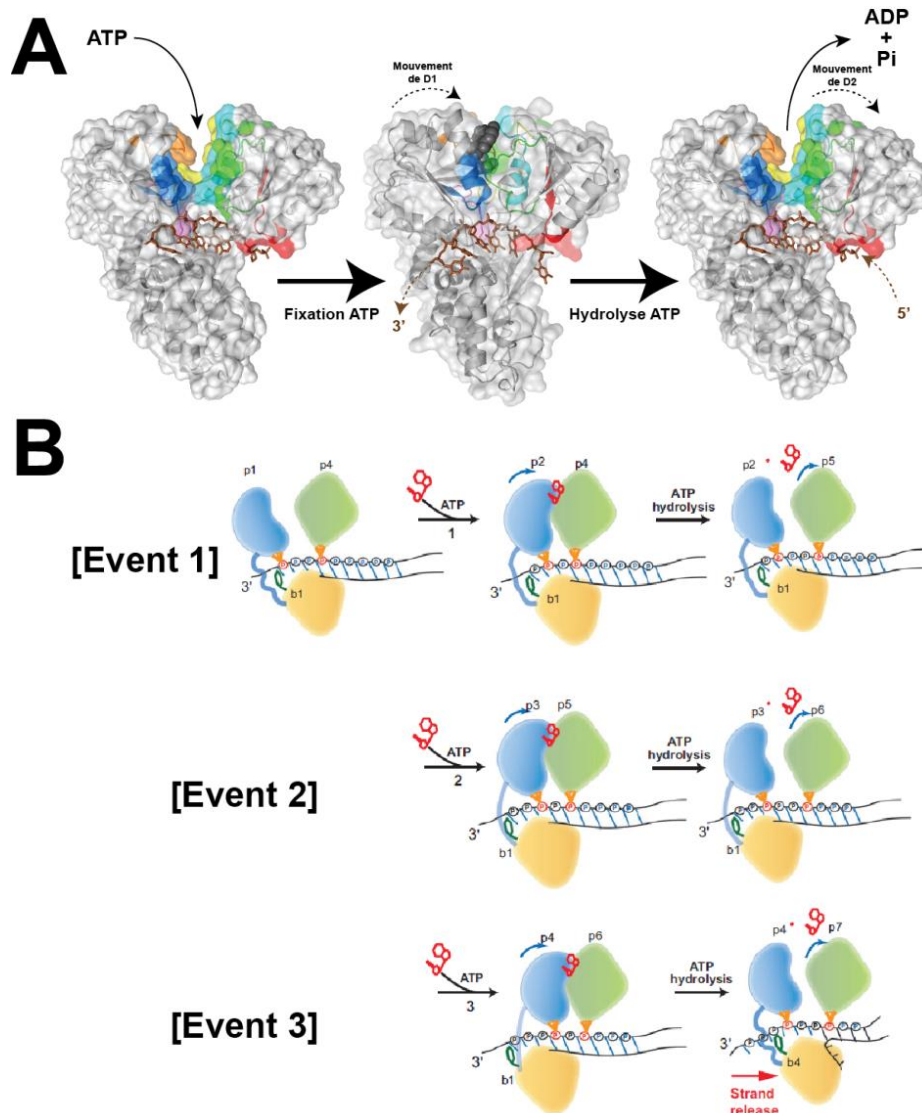


Figure 21 : Mécanisme de translocation de l'hélicase NS3. (A) La fixation d'une molécule d'ATP dans le cœur hélicase provoque la fermeture de la poche ATPasique via le mouvement de D1. L'hydrolyse (et le relargage) du nucléotide cofacteur provoque l'ouverture de la poche via le mouvement de D2. Cet événement s'accompagne de la translocation d'un nucléotide de la chaîne de l'ARN (PDB : 3KQK & 3KQL). (B) "Spring-loaded mechanism". L'hydrolyse d'ATP induit l'avancée 3'→5' des domaines D1 et D2 (bleu et vert respectivement) tandis que le domaine CTD (jaune) reste bloqué par la jonction double brin. L'énergie accumulée est libérée périodiquement (tous les 3 ATP hydrolysés) par l'avancée brusque du domaine CTD et par l'ouverture d'une partie du duplexe (par un mécanisme d'exclusion stérique) (adapté de Pyle, 2008)).

Récemment, un modèle similaire a été proposé pour l'hélicase Upf1 appartenant à la SF1 avec néanmoins quelques différences notables. En combinant des expériences de nano-manipulation à l'échelle de la molécule unique et des tests biochimiques, il a été montré que la protéine Upf1 est capable de se déplacer de manière directionnelle 5'→3' sur l'ARN ou

l'ADN simple brin avec une grande processivité (Fiorini *et al.*, 2015). Grâce à cette activité de translocation processive, Upf1 est capable d'ouvrir de longues (1.2 kB) structures en tige boucle ou encore de provoquer la dissociation de protéines Gp32 associées à un brin ADN (Fiorini *et al.*, 2015). Ces deux exemples illustrent les mécanismes distincts employés par les translocases SF1 et SF2 qu'il n'est pas possible de présenter plus en détails dans ce mémoire (pour revue: Ding & Pyle, 2012; Fairman-Williams *et al.*, 2010).

Ce couplage entre translocation unidirectionnelle et activité hélicase n'est pas universel. Certaines hélicases SF2 sont capables de translocation bipolaire ($5' \rightarrow 3'$ et $3' \rightarrow 5'$) *in vitro* (Prp43, par exemple) (Tanaka & Schwer, 2006; Tanaka *et al.*, 2007). Pour d'autres, la notion de translocation est même discutable (et discutée) dans la mesure où elles ne sont capables de dissocier que de très courts duplexes sans exhiber de polarité apparente : par exemple, elles peuvent, *in vitro*, dissocier des duplexes ARN à bouts francs même si leur efficacité est augmentée par la présence d'une extrémité sortante en 3' ou 5' (Bizebard *et al.*, 2004; Halls *et al.*, 2007; Rogers *et al.*, 2001; Yang & Jankowsky, 2006). C'est notamment le cas de spécimens de la famille DEAD-box (**Tableau 2 ; Figure 18 & Figure 19**), l'une des plus abondantes et physiologiquement importantes familles d'ARN hélicases. Certaines ARN hélicases DEAD-box sont capable de séparer les brins d'un court hybride ARN en présence d'analogues non-hydrolyzables de l'ATP, l'hydrolyse du nucléotide cofacteur n'étant nécessaire que pour le relargage de l'ARN et le recyclage de l'enzyme (**Figure 22**) (Liu *et al.*, 2008). Il est proposé que la seule présence du NTP à l'interface des domaines D1 et D2 stimule la fixation de l'ARN et provoque une distorsion locale de la double hélice ainsi que l'ouverture des paires de bases du duplexe (Chen *et al.*, 2008; Hilbert *et al.*, 2009; Mallam *et al.*, 2012). De par ces caractéristiques, les protéines DEAD-box ne sont généralement pas considérées comme des translocases mais plutôt comme des "pinces" (ou chaperonnes) ATP-dépendantes capables de provoquer la dissociation de courts duplexes de manière locale (Pyle, 2008; Redder *et al.*, 2015). Néanmoins, *in vivo*, les protéines DEAD-box sont associées à divers partenaires pouvant moduler leurs propriétés. Par exemple, l'hélicase eIF4A qui, seule, ne peut dissocier que de courts duplexes de manière non processive (Rogers *et al.*, 1999) est capable, en présence de ses partenaires eIF4H/B et eIF4G, d'une translocation $5' \rightarrow 3'$ le long de l'ARN suffisamment processive pour provoquer l'ouverture d'une structure tige-boucle de 72 pb (García-García *et al.*, 2015). Cette découverte récente pourrait marquer un changement de paradigme pour cette importante classe d'enzymes.

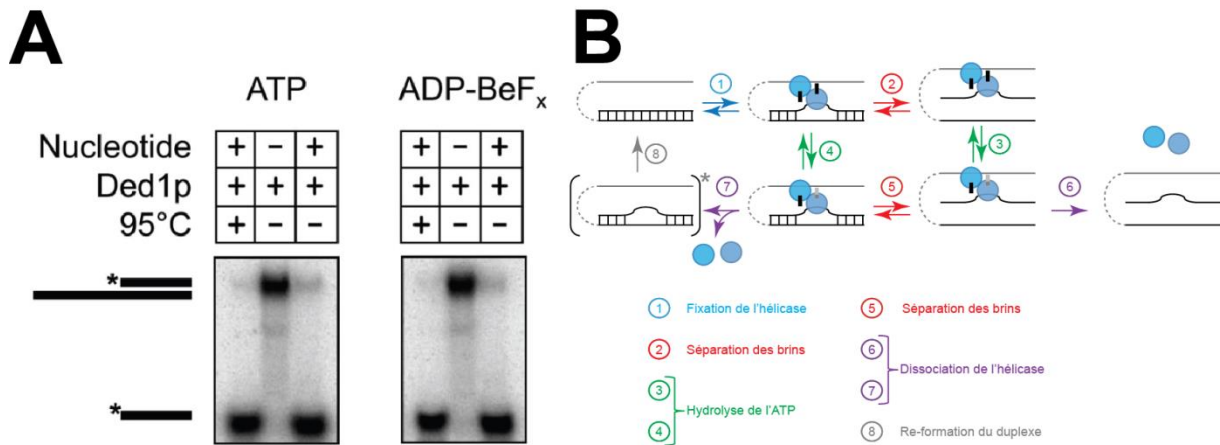


Figure 22 : Mécanisme d'action des DEAD-box hélicases. (A) La protéine DEAD-box Ded1p est capable de provoquer la séparation d'un court duplexe (13 bp) ARN:ARN en présence d'ATP ou d'un analogue non hydrolysable (ADP-BeF_x) (d'après (Liu et al., 2008)). (B) Mécanisme d'action proposé pour les protéines DEAD-box. La présence de l'ATP (trait noir épais) stimule la fixation de l'enzyme sur son substrat ARN provoquant l'ouverture locale de paires de bases. L'hydrolyse de l'ATP (ADP représenté par un trait gris épais) provoque le relargage du brin ARN et le recyclage de l'enzyme. La formation de dimères DEAD-box, telle que représentée ici, reste à ce jour spéculative (adapté de Jankowsky, 2011).

C. Rôles et importance

Les ARN hélicases SF1 et SF2 sont retrouvées dans tous les règnes du vivant où elles semblent impliquées dans la plupart des étapes du métabolisme ARN (Anantharaman *et al.*, 2002). Leur nombre est généralement lié à la complexité de l'organisme : les génomes d'*E. coli*, de *Saccharomyces cerevisiae* et d'*Homo sapiens* codent pour 5, 26 et 37 protéines DEAD-box, respectivement (cf : www.rnahelicase.org). Cette différence peut s'expliquer par le fait que, au cours de l'évolution, les eucaryotes ont développé un métabolisme ARN plus complexe nécessitant un plus grand nombre de processus pour lesquels le repliement des ARN est essentiel (épissage, export des ARNm du noyau, ribogenèse plus complexe nécessitant un système 18S au lieu de 16S, etc.) (Iost *et al.*, 2013). A titre de comparaison, 19 ARN hélicases SF1 et SF2, la plupart essentielles, sont impliquées dans la ribogenèse chez *S. cerevisiae* (Martin *et al.*, 2013) alors que seules quatre spécimens DEAD-box sont impliquées dans ce processus chez *E. coli* (voir Chapitre I page 37) et peuvent être inactivées sans effet délétère dans des conditions standard de croissance (Jagessar & Jain, 2010). A basse température, en revanche, la présence de ces protéines DEAD-box devient importante pour la croissance bactérienne (Chamot & Owtrim, 2000; Charollais *et al.*, 2003; Charollais *et al.*, 2004; Lechnik-Habrink *et al.*, 2013; Lim *et al.*, 2000; Rosana *et al.*, 2012) et est même essentielle pour certaines espèces psychrotolérantes (Markkula *et al.*, 2012; Pandiani *et al.*, 2010). Dans ces conditions, les structures secondaires et tertiaires des ARN sont plus stables

(y compris les structures incorrectes) (Cartier *et al.*, 2010) et leur présence au sein des ARNm/ARNr peut induire un défaut de traduction (Broeze *et al.*, 1978). L'action des protéines DEAD-box, notamment comme ARN chaperonnes, réduirait la formation de ces structures délétères (Iost *et al.*, 2013; Kuhn, 2012).

On retrouve aussi des ARN hélicases SF1 et SF2 dans trois familles de virus : l'hélicase CI chez les *Potyviridae* (génome ARN⁺) (Eagles *et al.*, 1994), l'hélicase NPH-II chez les *Poxviridae* (génome ADN double brin) (Paoletti & Moss, 1974), et l'hélicase NS3 chez les *Flaviviridae* (génome ARN⁺) (Gorbalenya *et al.*, 1989). Ces trois protéines font partie de la même famille au sein de la SF2 (**Figure 18 & Tableau 2**) et ont fait l'objet de nombreuses études qui ne sont pas détaillées ici (Fairman-Williams & Jankowsky, 2012; Myong *et al.*, 2007). Aujourd'hui, le rôle de ces hélicases dans le cycle viral reste assez mal défini mais pourrait être associé à la transcription et/ou la réplication du génome viral ou encore à l'assemblage des particules virales et à leur propagation (Gross & Shuman, 1996; Raney *et al.*, 2010; Revers & García, 2015).

2. LES HÉLICASES TOROÏDALES

La majorité des hélicases, appartenant aux familles SF3, SF4, SF5 et SF6 possèdent une structure similaire en anneau hexamérique au centre duquel va circuler le substrat acide nucléique. Cette structure est formée par six protomères de la même protéine, s'articulant pour former un anneau (**Figure 23**). Néanmoins il existe quelques exceptions comme l'hélicase ADN eucaryotique MCM qui forme des hétéro-hexamères (et des structures dodécamériques en double anneau) composés des protéines MCM-4, MCM-6 et MCM-7 (Patel & Picha, 2000; You *et al.*, 1999). La configuration en anneau hexamérique est très répandue pour les ADN hélicases mais rare pour les ARN hélicases (Rabhi *et al.*, 2010b) pour lesquelles il n'existe que deux exemples bien documentés: le facteur Rho (SF5) et le moteur d'empaquetage génomique P4 (SF4) des phages ϕ (Ewen & Revel, 1990). A noter que l'antigène large du virus SV40 (SF3) est une ADN hélicase qui possède aussi une activité sur des substrats ARN *in vitro* (Scheffner *et al.*, 1989) mais dont la pertinence physiologique n'est pas établie (Rabhi *et al.*, 2010b). Nous nous contenterons donc d'évoquer les cas de Rho et de P4 comme "véritables" ARN hélicases hexamériques.

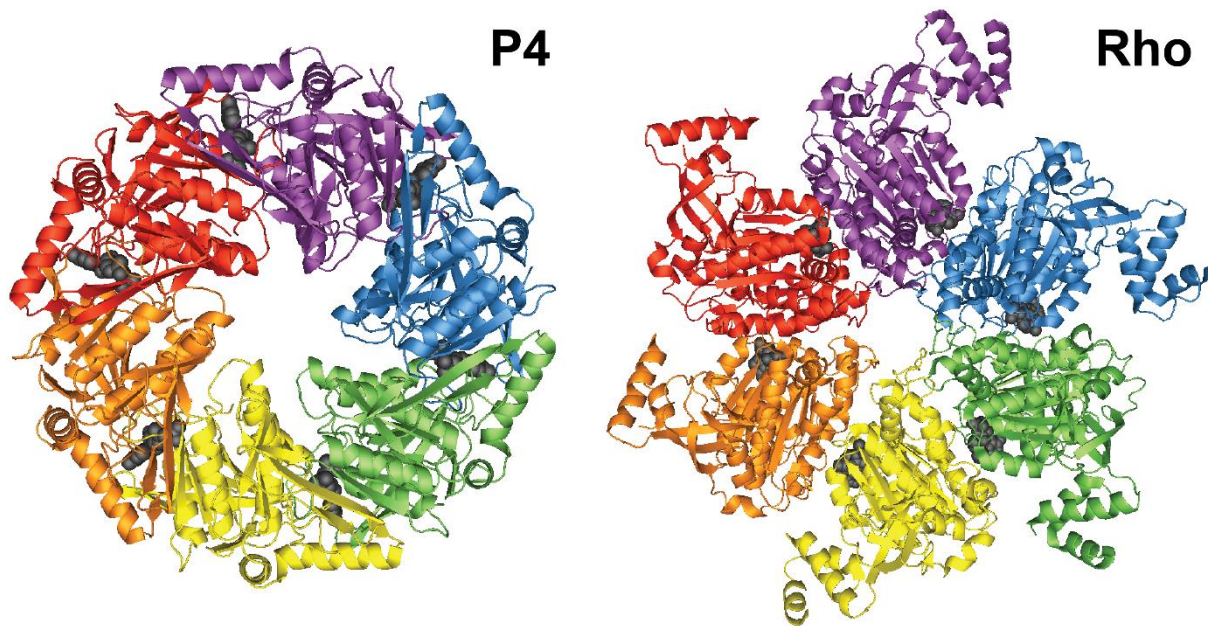


Figure 23 : Structure des ARN hélicases P4 et Rho. Rho et P4 sont les deux seules ARN hélicases dont la forme active s'organise en hexamère. Ici chaque sous-unité est représentée d'une couleur différente. Les analogues nucléotidiques (ADP pour P4 et ADP•BeF3 pour Rho) présents aux interfaces entre sous-unités sont représentés en gris. (P4 PBD : 1W44 ; Rho PBD : 3ICE). Le substrat ARN (non représenté) est censé circuler au centre de l'anneau hexamérique.

A. Organisation structurale

La conformation en anneau autour de l'acide nucléique apporte théoriquement plusieurs avantages, le principal étant qu'elle contribue à la processivité de l'enzyme (Patel & Picha, 2000; Rabhi *et al.*, 2010b). Le fait que l'acide nucléique soit piégé au centre de l'anneau est censé réduire le risque de dissociation du complexe lié au remodelage continu des interactions ARN:protéine intervenant durant la translocation (Rabhi *et al.*, 2010b). Il n'est donc pas surprenant de retrouver cette organisation chez des ADN hélicases impliquées dans la réplication du génome étant donné que ces enzymes doivent assurer la séparation de doubles brins sur de longues distances (Delagoutte & von Hippel, 2002; Remus & Diffley, 2009). Bien que les processivités de Rho et de P4 varient selon les espèces bactériennes ou les souches de phages chez qui elles sont isolées (Kainov *et al.*, 2003; Kalarickal *et al.*, 2010), elles restent généralement supérieures à celles mesurées pour nombre d'ARN hélicases monomériques (Jankowsky & Fairman-Williams, 2010). Un autre avantage de l'organisation en anneau oligomérique est qu'il offre une grande surface protéique pour des interactions avec des cofacteurs (Rabhi *et al.*, 2010b). Chez *E. coli*, le facteur Rho apparaît être au centre d'un réseau d'interactions protéine/protéine ayant pour effet soit d'augmenter (NusG) (Chalissery *et al.*, 2011; Mooney *et al.*, 2009b), soit de diminuer (YaeO ; Gutierrez *et al.*, 2007, Hfq ;

Rabhi *et al.*, 2011a, Psu ; Pani *et al.*, 2009) l'efficacité de terminaison de transcription. Pour P4 les principaux partenaires identifiés sont les protéines de la capsid qui peuvent augmenter la stabilité de l'hexamère (Lisal *et al.*, 2006) ou participer directement à l'action d'empaquetage du génome (Kainov *et al.*, 2006).

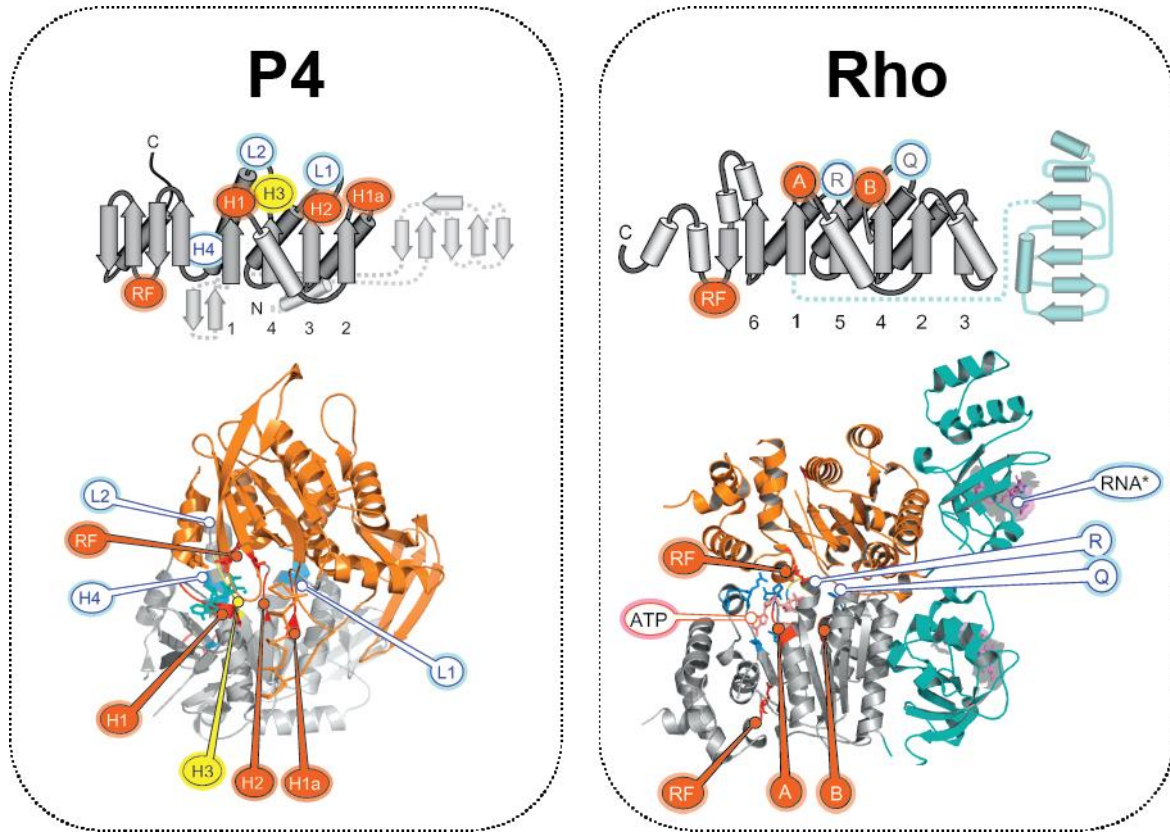


Figure 24 : Les domaines de type RecA de P4 et Rho forment des poches d'hydrolyse aux interfaces de chaque protomère. Pour chaque enzyme la topologie du domaine de type RecA est représentée en haut et les motifs importants sont mis en avant selon le code couleur suivant : orange, fixation et hydrolyse de l'ATP ; blanc, fixation de l'ARN ; jaune, communication entre l'hydrolyse de l'ATP et la fixation de l'ARN. En bas de chaque panneau est représentée la poche ATPase formée à l'interface de deux sous-unités (adapté de Fairman-Williams *et al.*, 2010).

Tout comme les ARN hélicases SF1 et SF2, les hélicases hexamériques possèdent aussi des repliements de type RecA, (ou de type apparenté AAA+ - **Figure 24**) (Berger, 2008) mais une différence importante est qu'il n'y a qu'un seul domaine de type RecA par protomère. Les poches d'interaction avec les NTP sont formées aux interfaces entre sous-unités (**Figure 23 & Figure 24**) pour un total de six poches NTPase par hexamère (Bogden *et al.*, 1999; Kainov *et al.*, 2008; Mancini *et al.*, 2004; Skordalakes & Berger, 2006; Thomsen & Berger, 2009). Au sein des domaines de type RecA, on retrouve des motifs équivalents à ceux présents chez les hélicases monomériques SF1/SF2 et/ou de nombreuses autres NTPases. C'est notamment le cas des motifs canoniques "Walker A" (H1 chez P4 - **Figure 24**) et "Walker B" (H1a et H2 chez P4 - **Figure 24**) (Walker *et al.*, 1982), correspondant aux motifs

I et II des hélicases SF1/SF2 (**Figure 19**). Le rôle du motif VI (stabilisation de l'état de transition de la réaction ATPase) des hélicases SF1/SF2 est quant à lui supporté chez Rho ou P4 par un "doigt arginine" situé sur la sous-unité adjacente (RF sur la **Figure 24**). En revanche, les motifs de Rho ou P4 impliqués dans la fixation et la translocation de l'ARN n'ont pas d'équivalent chez les hélicases SF1/SF2. Ces motifs, appelés Q-Loop et R-Loop (à ne pas confondre avec les R-loops ARN:ADN; voir page 24) pour Rho et L1 et L2 pour P4 (**Figure 24**), sont projetés (à partir de chaque protomère) au centre de l'hexamère où ils constituent un site d'interaction composite avec l'ARN qui s'y trouve. L'état d'hydrolyse des nucléotides présents dans les poches ATPase gouverne les conformations de ces boucles et leur capacité à interagir avec l'ARN. Bien que possédant de nombreux points communs avec les ADN hélicases toroïdales, Rho et P4 restent des enzymes strictement ARN-spécifiques, incapables de transloquer l'ADN. Ceci peut probablement s'expliquer par des contacts spécifiques formés entre les boucles d'interaction (e.g. Q-Loop) et les groupements 2'-OH du squelette carboné de l'ARN, comme cela a été démontré pour Rho (Thomsen & Berger, 2009).

Rho et P4 possèdent également des domaines supplémentaires leur conférant des particularités qui sont spécifiques à chacune. Chez Rho ces domaines sont situés du côté NTD de l'enzyme et sont impliqués dans la reconnaissance initiale et la fixation de l'enzyme à son substrat ARN (voir page 73) (Canals *et al.*, 2010; Skordalakes & Berger, 2003), tandis que chez P4 ces domaines sont situés du côté CTD et sont requis pour l'interaction avec les autres éléments de la capsid virale (Lisal *et al.*, 2006).

B. Mécanismes généraux

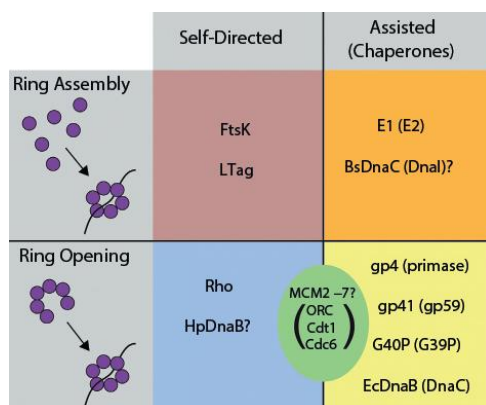


Figure 25 : Modèles de formation des complexes entre une hélicase toroïdale et son substrat acide nucléique (adapté de O'Shea & Berger, 2014).

L'organisation structurale des hélicases hexamériques requiert *a priori* un mécanisme de fonctionnement plus complexe que pour les hélicases SF1/SF2. Par exemple, la formation

d'un complexe enzyme:substrat actif nécessite que l'acide nucléique entre et occupe le centre de l'anneau. Cette configuration peut en fait résulter de deux modes d'assemblage distincts (**Figure 25**): i) assemblage direct (*ring assembly*) où l'hexamère est formé par la fixation extemporanée des six protomères directement sur l'acide nucléique ou ii) ouverture transitoire de l'anneau préformé (*ring opening*) qui nécessite la rupture des contacts entre deux sous-unités pour permettre l'entrée de l'acide nucléique au centre de l'hexamère (O'Shea & Berger, 2014). Cette deuxième stratégie est celle utilisée par le facteur Rho (voir page 73 - **Figure 25**). Les deux modes d'assemblage peuvent nécessiter l'assistance de facteurs auxiliaires (**Figure 25**) ou la reconnaissance d'une séquence spécifique au sein de l'acide nucléique (O'Shea & Berger, 2014), cette reconnaissance étant généralement assurée par un domaine accessoire dédié (Patel & Picha, 2000). Un mode d'assemblage alternatif repose sur l'entrée directe du substrat au centre de l'hexamère (comme un fil dans une tête d'aiguille). Ce dernier implique que le substrat soit obligatoirement sous une forme linéaire et que l'une des extrémités soit disponible, expliquant pourquoi il est proposé principalement pour les moteurs d'empaquetage de génomes viraux tels que P4, pour laquelle la présence des autres protéines de la capside facilite l'entrée de l'ARN au centre de l'anneau (Kainov *et al.*, 2003).

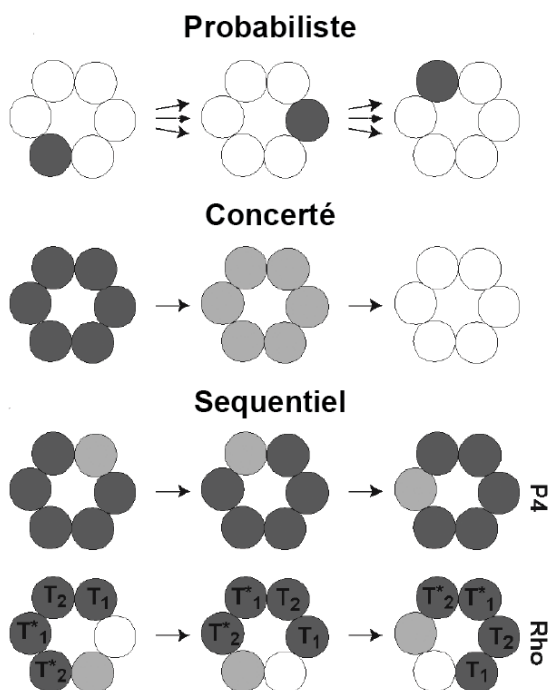


Figure 26 : Mécanismes d'hydrolyse des NTP proposé pour les hélicases hexamériques.

Les sous-unités blanches possèdent des sites d'hydrolyses vides, les sous unités gris clair des sites d'hydrolyses occupés par des produits d'hydrolyse (NDP+Pi) et les sous unités gris foncé des sites d'hydrolyses occupés par des NTP (adapté de Rabhi *et al.*, 2010b).

Une fois fixées sur leur substrat, les hélicases hexamériques vont pouvoir se transloquer en utilisant l'énergie issue de l'hydrolyse des NTP. La présence de six poches ATPase potentiellement actives présuppose différents mécanismes d'hydrolyse possibles. Il a été proposé que l'antigène large du virus SV40 utilise un mécanisme d'hydrolyse concertée

(**Figure 26**), où les six poches catalytiques fixent et hydrolyse les NTP en même temps (Gai *et al.*, 2004). Un autre mécanisme envisageable (mais pour lequel il n'existe pas d'exemple d'hélicase toroïdale connu) est que l'hydrolyse des NTP se fasse de manière aléatoire au sein des six sous unités de l'hexamère (**Figure 26**, modèle probabiliste). Enfin, Rho et P4 utilisent toutes les deux un mécanisme dit séquentiel (**Figure 26**) où les événements d'hydrolyses ont lieu les uns à la suite des autres de manière rotative au sein de l'hexamère (Rabhi *et al.*, 2010b).

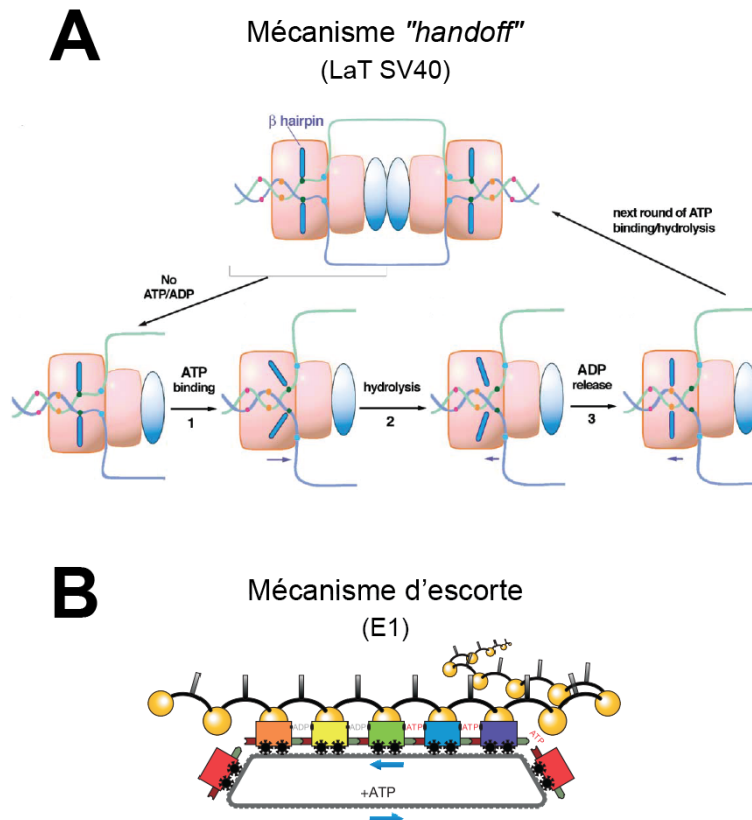


Figure 27 : Mécanismes de translocation utilisés par les hélicases hexamériques

(A) L'antigène large du virus SV40 est proposé se transloquer selon un modèle *handoff*. Selon l'état hydrolytique des poches ATPase, les anneaux adoptent des conformations différentes supposées transloquer le brin ADN au travers de l'anneau (adapté de (Gai *et al.*, 2004)). (B) L'hélicase E1 est proposée utiliser un mécanisme dit d'escorte où chaque sous unité (représentée par un wagonnet) prend en charge un nucléotide et l'accompagne au travers l'anneau suivant l'état hydrolytique de la poche ATPasique. (d'après Enemark & Joshua-Tor, 2006).

La capacité des hélicases hexamériques à se transloquer sur l'acide nucléique est liée aux réarrangements structuraux au sein des protomères (et/ou au changement de disposition relative des protomères) découlant directement de l'état chimique des poches ATPase (occupées par une molécule NTP, occupées par un NTP et catalytiquement actives, occupées par les produits NDP+Pi, vides ou en état d'échange) (Enemark & Joshua-Tor, 2008). Selon le mécanisme d'hydrolyse utilisé par l'enzyme, le mécanisme de translocation peut varier. Par exemple, il est proposé pour l'antigène large du virus SV40 que la translocation de l'ADN se fait par un mécanisme de type "*handoff*" où l'hydrolyse concertée induit un changement conformationnel global permettant la traction de l'ADN au centre de l'hexamère par les six sous-unités à la fois (**Figure 27.A**) (Gai *et al.*, 2004). Pour l'ADN hélicase E1, qui utilise un mécanisme d'hydrolyse séquentiel, une structure cristalline suggère un mécanisme dit

d'escorte (Enemark & Joshua-Tor, 2006). Ce mécanisme stipule que chaque nucléotide est pris en charge par une sous-unité qui, au fur et à mesure de l'évolution chimique de la poche ATPase associée, va lui faire traverser le pore de l'anneau (**Figure 27.B**). Des mécanismes de translocation similaires à celui décrit pour E1 ont été proposés pour Rho et P4 (Kainov *et al.*, 2008; Rabhi *et al.*, 2010b; Thomsen & Berger, 2009).

In vitro, Rho et P4 montrent une activité hélicase avec des duplexes ARN:ARN ou ARN:ADN à condition que ces duplexes soient précédés par une région 5'-ARN simple-brin non structurée (Kainov *et al.*, 2003; Walmacq *et al.*, 2004). Cette "extension" est requise pour le recrutement et la fixation par l'enzyme. Dans le cas de Rho, l'extension simple-brin doit contenir une séquence *Rut* (voir page 81) (Walmacq *et al.*, 2004; Walmacq *et al.*, 2006) alors que seule la longueur de l'extension (une centaine de nucléotides) est déterminante pour P4 (la reconnaissance de l'ARN cible étant supportée par d'autres protéines de la capsid *in vivo*; (Kainov *et al.*, 2003). L'importance de la position de l'extension simple-brin, en amont de la région duplexe, vient du fait que Rho et P4 sont toutes deux des ARN translocases unidirectionnelles 5'→3' (Rabhi *et al.*, 2010b). Néanmoins, les deux enzymes ont une disposition inversée de leur anneau sur la chaîne ARN (face NTD→3' pour Rho et face CTD→3') comme si l'une utilisait la "marche avant" et l'autre la "marche arrière" (**Figure 28**) (Rabhi *et al.*, 2010b). Cette inversion est le reflet de la différence de mécanismes de translocation utilisés par ces enzymes issus de superfamilles différentes (Kainov *et al.*, 2008; Thomsen & Berger, 2009).

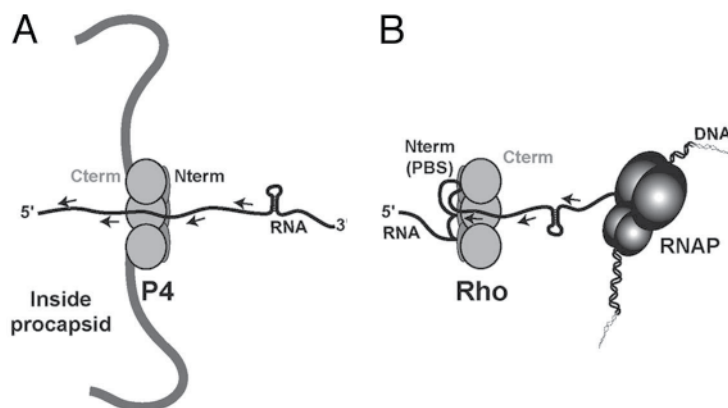


Figure 28: Actions de Rho et P4 dans leurs contextes biologiques.

La translocation 5'→3' de P4 est utilisée pour empaqueter le génome viral lors du cycle lytique du phage. Pour Rho, cette translocation permet à l'enzyme de rattraper l'ARNP et de provoquer la terminaison de la transcription. (D'après Rabhi *et al.*, 2010b)

La façon la plus probable permettant à Rho et P4 de provoquer la séparation des duplexes ARN/ARN (ou ARN/ADN) repose sur un principe d'exclusion stérique: la translocation de l'anneau le long de l'ARN entraîne l'élimination d'obstacles (duplexe ou complexe nucléoprotéique) car son pore est trop petit pour les laisser passer et ne peut qu'accommoder la chaîne ARN simple-brin (Patel & Picha, 2000). Dans le cas de Rho, ce

modèle est supporté par des données biochimiques et structurales (voir Chapitre III) (Skordalakes & Berger, 2006; Thomsen & Berger, 2009; Walmacq *et al.*, 2006). Ce modèle est également compatible avec les différents modèles structuraux de P4 (Kainov *et al.*, 2008; Mancini *et al.*, 2004) même s'il n'a pas été directement testé pour cette enzyme.

Chapitre III : Le facteur Rho

Comme nous avons pu le voir précédemment, le facteur Rho est un acteur central du métabolisme bactérien (Chapitre I). De plus, c'est un exemple unique d'ARN hélicase de par son organisation structurale et son mécanisme de fonctionnement "mécanochimique" (Chapitre II). Dans ce chapitre, j'approfondis ces notions en détaillant les aspects pertinents de la structure, des activités enzymatiques, et des fonctions physiologiques de Rho qui ont été étudiées principalement chez *E. coli*. Je conclus ce chapitre en abordant ce que l'on sait de l'importance du facteur Rho au sein de la biodiversité.

1. ORGANISATION STRUCTURALE

La première séquence de Rho disponible, fut la séquence de la protéine d'*E. coli* (EcRho) (Pinkham & Platt, 1983). Des expériences de protéolyse ménagée ont montré que chaque sous-unité d'EcRho était composée de deux domaines (NTD et CTD) structuralement distincts (Bear *et al.*, 1985; Dombroski & Platt, 1988). Le NTD est composé des 130 premiers acides aminés d'EcRho et contient deux sous-domaines – le sous-domaine NHB (*N-terminal Helix Bundle*) et un sous-domaine de type CSD (*Cold Shock Domain*, contenant les motifs clés RNP1 et RNP2) – (**Figure 29.A**) impliqués dans le recrutement du brin ARN et la reconnaissance des sites *Rut* (Bogden *et al.*, 1999; Canals *et al.*, 2010; Skordalakes & Berger, 2003). Le CTD, quant à lui, comprend les motifs essentiels à l'activité de moteur moléculaire d'EcRho, c'est à dire *i*) les motifs Walker A (P-Loop) et Walker B et les résidus "doigt arginine", "glutamate catalytique" et "valve arginine" (**Figure 29.A**) impliqués dans la fixation et l'hydrolyse de l'ATP et *ii*) les motifs Q-Loop et R-Loop impliqués dans la fixation et la translocation de la chaîne ARN par EcRho (Chalissery *et al.*, 2007; Miwa *et al.*, 1995; Thomsen & Berger, 2009).

Dès sa découverte, il a été suggéré qu'EcRho adopte une configuration oligomérique pour exercer son activité (Roberts, 1969). Diverses caractérisations d'EcRho par microscopie électronique (Gogol *et al.*, 1991; Yu *et al.*, 2000), ultracentrifugation analytique (Geiselman *et al.*, 1992) ou cristallographie aux rayons X (Skordalakes & Berger, 2003; Skordalakes & Berger, 2006; Thomsen & Berger, 2009) ont plus tard démontré que le facteur s'organise en un anneau homo-hexamérique (**Figure 29.B**). De plus, les études par microscopie électronique et cristallographie ont indiqué que l'hexamère EcRho pouvait adopter une configuration ouverte ou fermée suivant les conditions expérimentales, illustrant probablement des états fonctionnels différents (voir ci-dessous). Ces caractéristiques ont été

largement retrouvées lors d'une l'étude du facteur Rho de *Thermotoga maritima* (TmRho) par cristallographie aux rayons X (Canals *et al.*, 2010).

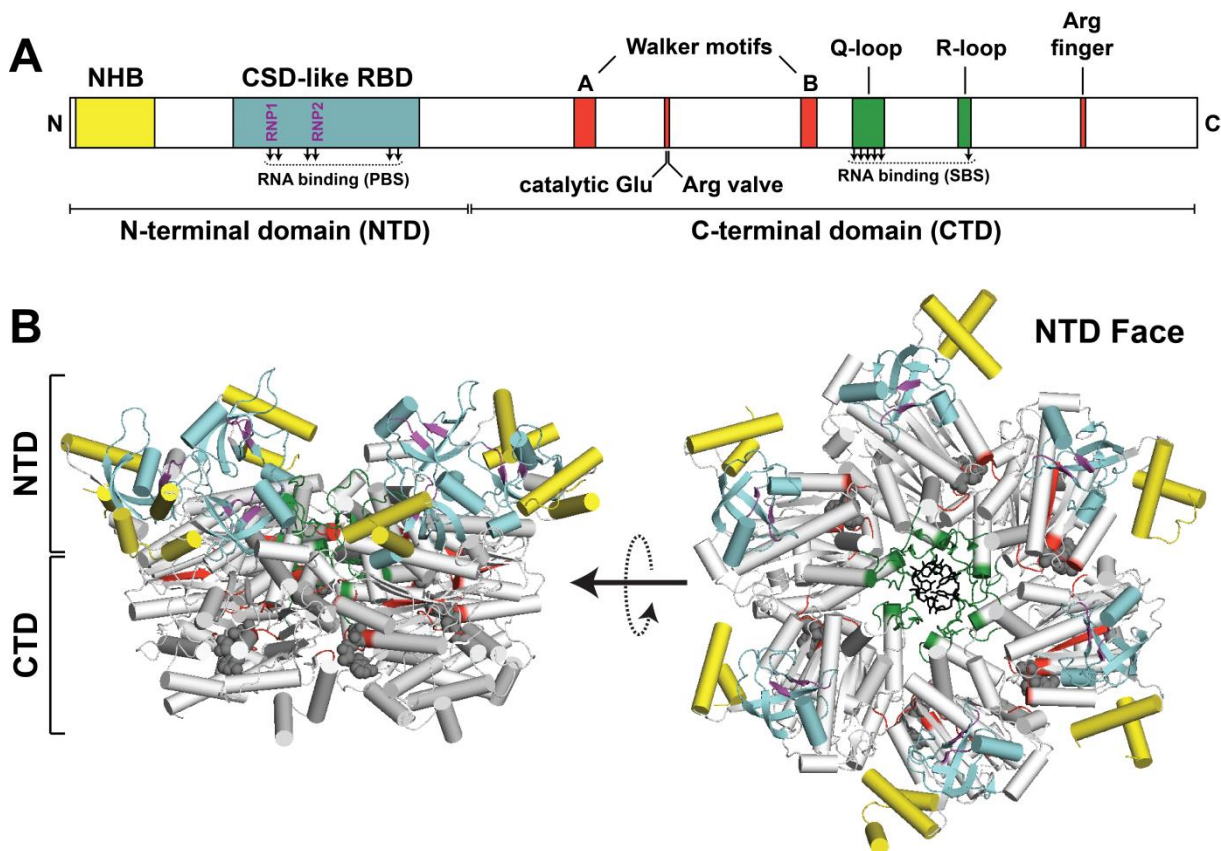


Figure 29 : Organisation structurale du facteur Rho. (A) Structure primaire de la protéine. Les différents sous-domaines des NTD et CTD sont représentés de couleurs différentes selon leurs fonctions : le recrutement (électrostatique) initial de l'ARN en jaune, la reconnaissance et la fixation de l'ARN en cyan, la fixation et l'hydrolyse du NTP en rouge, la fixation et la translocation le long de l'ARN en vert. Les flèches représentent les résidus en contact avec l'ARN dans les structures d'EcRho. (B) Structure de l'hexamère Rho (PDB: 3ICE) Les différents motifs sont colorés comme dans le panneau A. Les motifs RNP1 et RNP2 constituant les 6 poches de fixation spécifiques de l'ARN et formant le PBS sont représentés en violet. Les ADP-BeF₃ présents dans les poches ATPase sont représentés en gris (sphères) et L'oligonucléotide rU₆ est montré en noir, au centre de l'hexamère (d'après D'Heygere *et al.*, 2013).

Les différentes structures cristallines d'EcRho ont confirmé l'hypothèse ancienne (Richardson, 1982b) que ce facteur possède deux sites distincts de fixation à l'ARN (Bogden *et al.*, 1999; Skordalakes & Berger, 2003; Skordalakes & Berger, 2006; Thomsen & Berger, 2009). Le site primaire de fixation à l'ARN (PBS, pour *Primary Binding Site*) se situe sur la face NTD (**Figure 29**), tandis que le site secondaire d'interaction (SBS, pour *Secondary Binding Site*) se situe au centre de l'hexamère (**Figure 29**). Ces structures ont par ailleurs permis de déterminer que les sites d'hydrolyse de l'ATP (poches ATPase) sont situés aux interfaces entre protomères (**Figure 29**).

A. Le site primaire de fixation à l'ARN (PBS)

Rho possède une forte affinité pour les ADN et ARN simples brins riches en (désoxy)cytidines, le substrat préférentiel étant le poly[rC] (Lowery & Richardson, 1977; Lowery-Goldhammer & Richardson, 1974; Wang & von Hippel, 1993a; Wang & von Hippel, 1993b). Plusieurs structures cristallines d'EcRho en association avec des ligands ARN ou ADN ont permis d'expliquer cette préférence (Bogden *et al.*, 1999; Canals *et al.*, 2010; Skordalakes & Berger, 2003). Ainsi, les résidus Phe64, Arg66, Asp78, Tyr80, Glu108, et Tyr110 présents dans le sous-domaine CSD d'EcRho (**Figure 29.A**) forment une poche de reconnaissance spécifique pour un dimère de pyrimidines 5'-YC (Bogden *et al.*, 1999). L'occupation de ces poches par des dimères 5'-YC au sein de l'hexamère EcRho (**Figure 30**) constitue le seul type d'interaction Rho:ARN résolue cristallographiquement pour le PBS (Bogden *et al.*, 1999; Skordalakes & Berger, 2003; Skordalakes & Berger, 2006) même si des expériences biochimiques et de RMN suggèrent que le PBS comprend d'autres éléments (Briercheck *et al.*, 1998; Schwartz *et al.*, 2007b). Néanmoins, la spécificité des poches PBS pour les dimères 5'-YC explique probablement en grande partie la préférence d'EcRho pour les séquences riches en C (voir également la description des sites *Rut* ci-dessous; page 81). De plus, l'absence de contacts avec les groupements 2'-OH des dimères 5'-YC permet de comprendre pourquoi EcRho a des affinités comparables pour des ARN et des ADN simples brins de mêmes compositions (Richardson, 1982b; Richardson & Richardson, 1992).

Dans la première structure cristalline d'EcRho sous forme hexamérique, l'anneau adopte une configuration ouverte et contacte un dimère 5'-YC sur 5 de ses 6 protomères (Skordalakes & Berger, 2003). Les auteurs émettent l'hypothèse que les 6 poches de reconnaissance 5'-YC ensemble constituent le PBS, formant une sorte de couronne sur la surface NTD de l'hexamère (**Figure 30**). Dans une configuration ouverte de l'hexamère, cette disposition du PBS orienterait et faciliterait l'entrée du brin ARN dans le canal central de l'enzyme (Skordalakes & Berger, 2003). Dans une étude plus récente, l'hexamère du facteur TmRho adopte une configuration ouverte identique bien que la structure cristalline ait été obtenue en absence de ligand ARN ou ADN (Canals *et al.*, 2010). Les auteurs remarquent par ailleurs que le sous-domaine NHB présente une forte charge positive et apparaît être flexible. Ils proposent que le NHB est impliqué dans le recrutement électrostatique de l'ARN et dans sa présentation aux poches 5'-YC pour une reconnaissance plus spécifique. Bien que les deux études soient en désaccord sur le sens probable de circulation de la chaîne ARN sur la

couronne PBS, elles s'accordent sur le fait que la fixation de l'ARN au PBS oriente la région aval de l'ARN (en direction de l'extrémité 3'-OH) dans le canal central de l'hexamère où elle sera alors en mesure de contacter le SBS (**Figure 30**) (Canals *et al.*, 2010; Skordalakes & Berger, 2003).

Une question longtemps débattue était de savoir si l'ARN reste fixé au PBS durant la translocation de l'enzyme le long de la région aval de l'ARN. C'est ce que propose le modèle de “*tethered tracking*” (**Figure 30**) envisagé pour la première fois en 1994 à la suite d'expériences biochimiques reposant sur l'activité hélicase d'EcRho (Steinmetz & Platt, 1994). Ce modèle n'a été solidement validé que très récemment par des expériences de nano-manipulation d'EcRho à l'échelle de la molécule unique (Gocheva *et al.*, 2015; Koslover *et al.*, 2012) et par des expériences d'interférence chimique reposant sur la méthode trNAIP (*time-resolved Nucleotide Analog Interference Probing*) (Soares *et al.*, 2014).

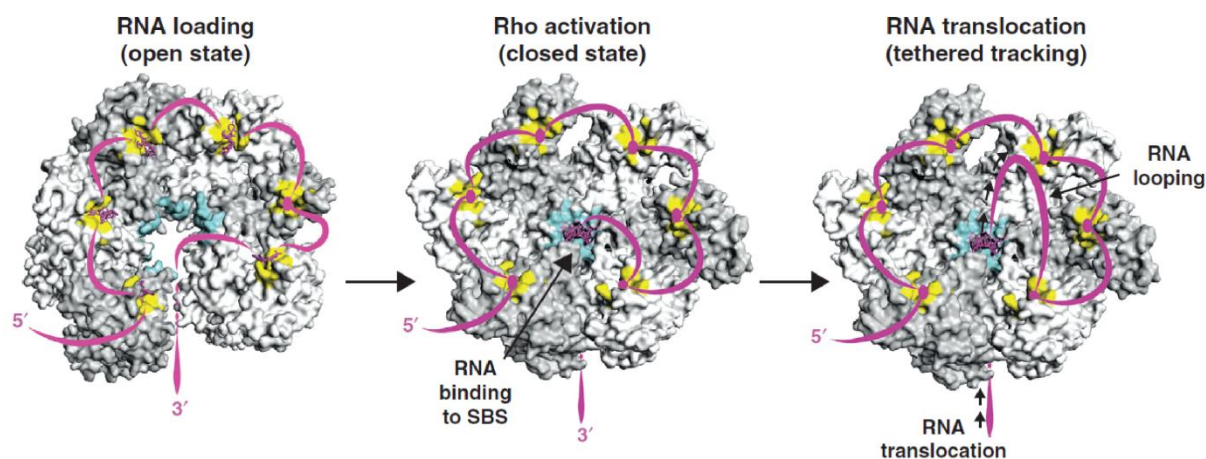


Figure 30 : Mécanisme de formation et d'activation du complexe Rho:ARN. La fixation de l'ARN (rose) sur les six sites constituant le PBS (jaune) favorise son entrée au centre de l'anneau, son contact avec le SBS (bleu clair) et la fermeture du complexe. La fixation de l'ARN sur le PBS perdure durant la translocation de l'enzyme, conduisant à l'extension d'une boucle ARN (RNA looping) sur la face NTD de l'enzyme. La circulation de l'ARN sur le PBS (ici dans le sens des aiguilles d'une montre) n'est pas formellement définie et reste débattue (d'après Boudvillain *et al.*, 2013)

B. Le site secondaire de fixation à l'ARN (SBS)

Bien qu'EcRho fixe aussi bien l'ADN que l'ARN (de mêmes séquences et riches en C) via son PBS, ses activités enzymatiques sont strictement ARN-dépendantes (Schwartz *et al.*, 2009; Wang & von Hippel, 1993b). Ceci peut être expliqué par la présence d'un second site de fixation (SBS) au sein de l'anneau hexamérique qui est strictement spécifique pour l'ARN (Richardson, 1982a; Wei & Richardson, 2001).

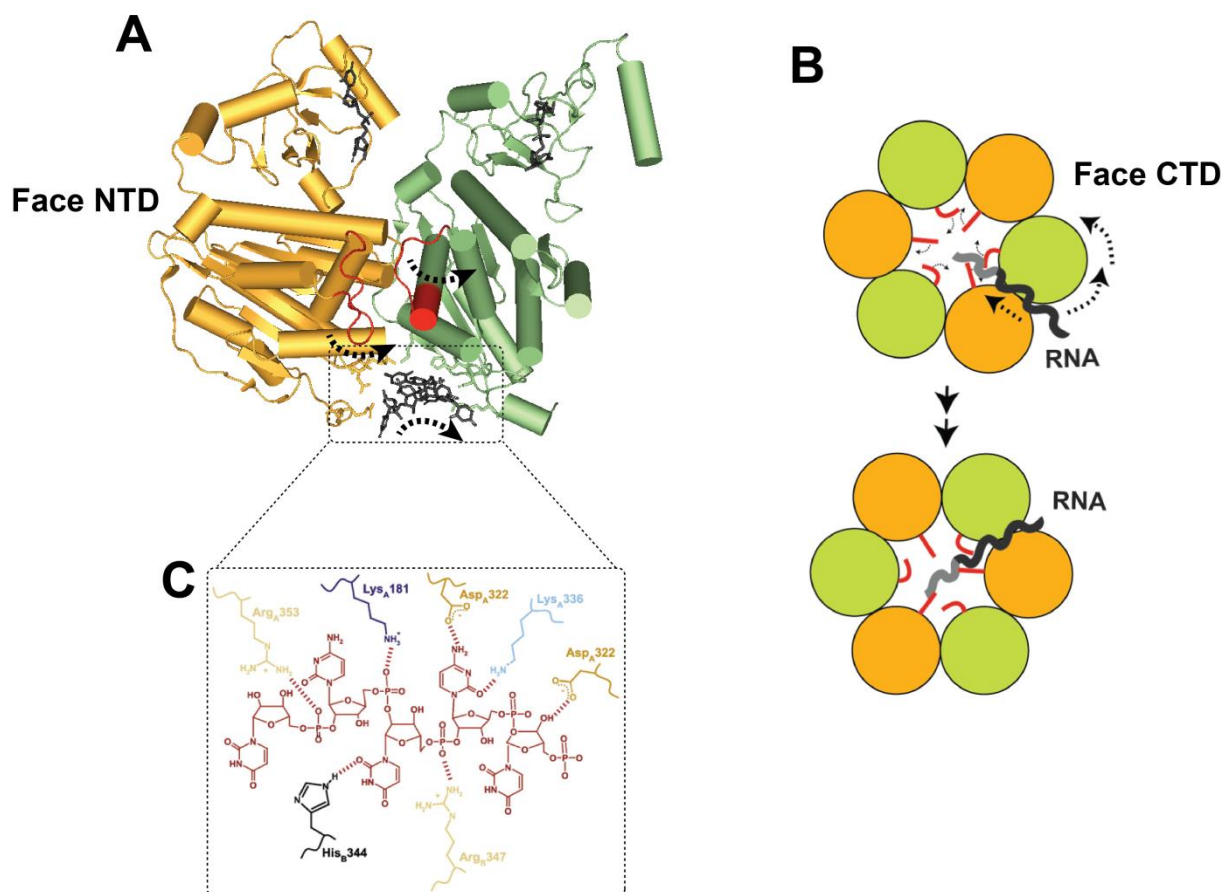


Figure 31 : Modèle de translocation par “RNA handoff”. (A) Structure d’un “dimère” EcRho formé à la jonction d’une maille cristalline (PDB : 2HT1). Les deux protomères sont représentés en vert et orange. Les motifs Q-Loop (rouge) adoptent des conformations différentes suivant le protomère. (B) Principe du modèle de “RNA handoff” (voir texte). (C) Réseau d’interactions avec le pentanucléotide 5’-UCUCU observées à l’interface CTD entre les deux protomères (adapté de Skordalakes & Berger, 2006).

En 2006, une première structure de l’hexamère EcRho en conformation fermée est publiée (Skordalakes & Berger, 2006). Dans cette structure, dite en “trimère de dimères” (organisée aux interfaces entre 3 mailles cristallines), différents fragments d’ARN sont résolus. Ainsi, au sein de chaque dimère, la crevasse formée sur la face CTD à l’interface entre protomères est occupée par un pentanucléotide 5’-UCUCU (**Figure 31.A**). Ce pentanucléotide forme des contacts avec des résidus importants et conservés comme la Lys181 (motif P-Loop) ou l’Asp322 (motif R-Loop) susceptibles de composer le SBS (**Figure 31.C**). Il est intéressant de noter que ce réseau SBS-ARN putatif est juxtaposé à la poche ATPase, ce qui suggère une communication directe entre celle-ci et le SBS notamment grâce à la Lys181. Un autre élément intéressant de la structure est la disposition des motifs Q-Loop qui adoptent alternativement (un monomère sur deux) une conformation “étendue” ou “repliée” au centre de l’anneau hexamérique (**Figure 31**). En s’appuyant sur ces observations, les auteurs proposent un modèle de translocation, dit de “RNA handoff”, basé sur l’interconversion ATP-dépendante entre deux groupes de sous-unités au sein de l’hexamère ;

(**Figure 31.B**) (Skordalakes & Berger, 2006). Cette inter-conversion mobiliserait les deux groupes de motifs Q-Loop correspondants qui, successivement, prendraient en charge la chaîne ARN pour la transloquer au travers le canal central de l'hexamère (en agissant comme des leviers) tout en la transférant à l'interface CTD directement adjacente (**Figure 31.B**). On peut noter que dans cette structure en “*trimère de dimères*”, des dinucléotides sont visibles (résolus) dans les poches 5'-YC du PBS tandis que des ions sulfates occupent les poches ATPase (Skordalakes & Berger, 2006). Autrement dit, cette structure est compatible avec un mode de translocation par “*tethered tracking*” (**Figure 30**) mais, bien que supportant un mécanisme d'hydrolyse concertée (voir Chapitre II page 64), ne permet pas d'évaluer précisément comment le processus d'inter-conversion entre monomères peut être couplé au cycle ATPase.

En 2009, une seconde structure de l'hexamère EcRho en conformation fermée est proposée par la même équipe (Thomsen & Berger, 2009). Cette structure ne présente pas d'ARN dans les poches 5'-YC du PBS mais contient un ligand (rU)₆ résolu au centre de l'anneau ainsi que des analogues nucléotidiques ADP•BeF₃ dans les poches ATPase (**Figure 29 & Figure 32**). De plus, les six protomères EcRho possèdent des conformations différentes et sont organisés de façon hélicoïdale, donnant un anneau asymétrique (**Figure 32.A**). Cette configuration est similaire à celle observée pour l'ADN hélicase E1 (Enemark & Joshua-Tor, 2006). Cinq des six sous-unités EcRho forment des contacts directs avec le ligand (rU)₆, chacune préférentiellement avec un des résidus uraciles. Ce réseau d'interactions SBS-ARN est complètement différent de celui observé pour la structure en “*trimère de dimères*”. Ici, l'ARN est contacté par des résidus des motifs Q-Loop (Val284, Thr286 et Gly287) ainsi que par les Lys326 des motifs R-Loop. Les squelettes amide des résidus Val284 contactent les groupements 2'-OH du ligand (rU)₆, ce qui pourrait expliquer la stricte spécificité de Rho pour l'ARN (Thomsen & Berger, 2009). En s'appuyant sur ces observations, les auteurs proposent un modèle de translocation alternatif, inspiré de l'ADN hélicase E1 (Enemark & Joshua-Tor, 2006), dit modèle d'escorte (**Figure 27 & Figure 32.C**). Selon ce modèle, chaque sous-unité prend successivement en charge un nucléotide et l'escorte au travers le canal central de l'hexamère. Cette translocation est assurée par les mouvements des motifs Q-Loop et R-Loop faisant suite aux réarrangements structuraux induits par la fixation et l'hydrolyse de l'ATP (**Figure 32**). Dans ce modèle, l'hydrolyse des molécules d'ATP au sein de l'hexamère est séquentielle (voir Chapitre II page 65) et le rendement mécano-chimique est d'une base d'ARN transloquée par molécule d'ATP hydrolysée (Thomsen & Berger, 2009).

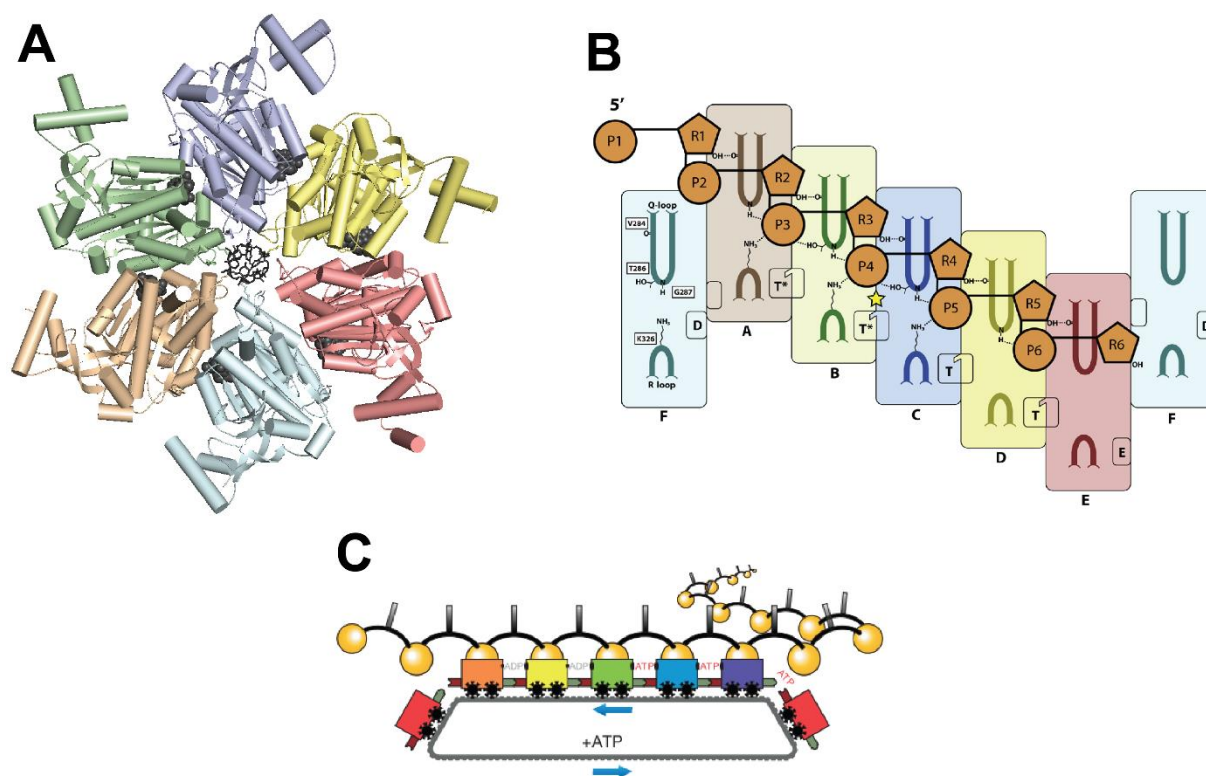


Figure 32 : Modèle de translocation par “RNA escort”. (A) Structure de l’hexamère EcRho organisé en anneau asymétrique (PDB: 3ICE). Les ADP•BeF₃ présents dans les poches ATPase sont représentés en gris et L’oligonucléotide rU₆ en noir, au centre de l’hexamère. Chaque sous-unité est d’une couleur différente. (B) Représentation schématique des contacts entre Rho et l’ARN en fonction des états chimiques des poches ATPase. Les monomères sont représentés “dans le plan” par des rectangles colorés comme dans le panneau A. Le monomère F est représenté deux fois afin de mettre en avant le décalage de sa position par rapport aux sous-unités A et E. Seules les motifs Q-Loop et R-Loop sont représentés sur chaque monomère ainsi que leurs résidus contactant l’ARN. Les rectangles aux interfaces représentent les poches ATPase et les expansions crantées la projection des doigts arginines. E, T, T* et D correspondent aux différents états d’occupation des poches ATPase (vide, ATP, [ATP]‡, ADP+Pi) décrits dans le paragraphe suivant. L’ARN est représenté par des ronds correspondant aux phosphates (P), et des pentagones correspondant aux riboses (R). Les interactions entre les sous-unités et les nucléotides sont représentées par des pointillés. L’étoile indique l’interface pour laquelle le plus d’interactions sont observables. (d’après Thomsen & Berger, 2009) (C) Modèle d’escort proposé pour l’ADN hélicase E1. Le code couleur est différent de celui des panneaux A et B (d’après Enemark & Joshua-Tor, 2006).

Malgré leurs mérites respectifs, les deux structures d’EcRho en anneau fermé (Skordalakes & Berger, 2006; Thomsen & Berger, 2009) suggèrent des mécanismes d’action radicalement différents. En 2009, une analyse des interactions formées entre Rho et l’ARN en cours de translocation par la technique NAIM (*Nucleotide Analog Interference Mapping*) révélait que Rho ne sonde pas tous les nucléotides de sa piste ARN de la même manière et ne requière un groupement 2’-OH que tous les ~ 7 nt en moyenne (Schwartz *et al.*, 2009). Cette observation semblait, à l’époque, plutôt en accord avec un mécanisme de translocation par “sauts” de ~ 7 nt, c’est-à-dire compatible avec l’action de “leviers” comme pour le “RNA handoff” (**Figure 31.B**). Néanmoins, la rupture de la seule interaction impliquant un groupement 2’-OH dans la structure en “trimère de dimères” (**Figure 31.C**) par la mutation

D322A s'est révélée sans effet délétère sur l'activité enzymatique d'EcRho (Schwartz *et al.*, 2009). Le manque de "représentativité" fonctionnelle de cette structure fut confirmé par une analyse plus large des résidus SBS putatifs par mutagenèse dirigée qui, en revanche, s'avèrera compatible avec l'organisation de l'hexamère EcRho en anneau asymétrique (**Figure 32**) (Rabhi *et al.*, 2011b). Par exemple, cette analyse, tout comme une simulation récente du mécanisme de translocation par dynamique moléculaire (Ma & Schulten, 2015), confirma que les résidus Lys326 (qui contactent la chaîne ARN seulement dans la structure asymétrique d'EcRho) jouent un rôle primordial.

C. Le site ATPase

Il fut très tôt observé qu'EcRho et le moteur F1-ATPase possèdent ~ 40% d'identité (Dombroski & Platt, 1988). Une analyse des séquences des deux protéines montre que Rho et la F1-ATPase partagent plusieurs motifs impliqués dans la fixation de l'ATP (Dombroski *et al.*, 1988; Dombroski & Platt, 1988). Cette similarité fut par la suite confirmée par la comparaison de leurs structures cristallines respectives (Skordalakes & Berger, 2006). Ainsi, les deux enzymes possèdent six poches ATPase localisées aux interfaces entre protomères (**Figure 33**) qui, à cet endroit, adoptent une organisation structurale ressemblant à la protéine RecA (on parle de domaines de type RecA) (Canals *et al.*, 2010; Skordalakes & Berger, 2003; Skordalakes *et al.*, 2005; Skordalakes & Berger, 2006; Thomsen & Berger, 2009). Ces poches sont constituées par les motifs Walker A et Walker B (Dombroski *et al.*, 1988; Dombroski & Platt, 1988), que l'on retrouve chez un grand nombre de protéines ATPase (**Figure 29**) (Walker *et al.*, 1982), et contiennent les résidus clés pour le mécanisme d'hydrolyse (**Figure 33**). Parmi eux, on retrouve le glutamate catalytique (Glu211), impliqué dans l'activation de la molécule d'H₂O attaquant le groupement γ -phosphoryle de l'ATP, la "valve arginine" (Arg212) qui détecte la présence ou l'absence du groupement γ -phosphoryle, et le "doigt arginine" (Arg366) qui contribue à stabiliser la charge négative qui se développe dans l'état de transition pentacoordinné de la réaction d'hydrolyse (Balasubramanian & Stitt, 2010; Thomsen & Berger, 2008). On retrouve également des résidus contribuant à stabiliser le complexe Mg•ATP dans la poche ATPase, comme la Phe355 (empilement de la purine) ou l'Asp265 (stabilisation indirecte de l'ion Mg²⁺ via une molécule d'eau) (**Figure 33**) (Balasubramanian & Stitt, 2010; Thomsen & Berger, 2009).

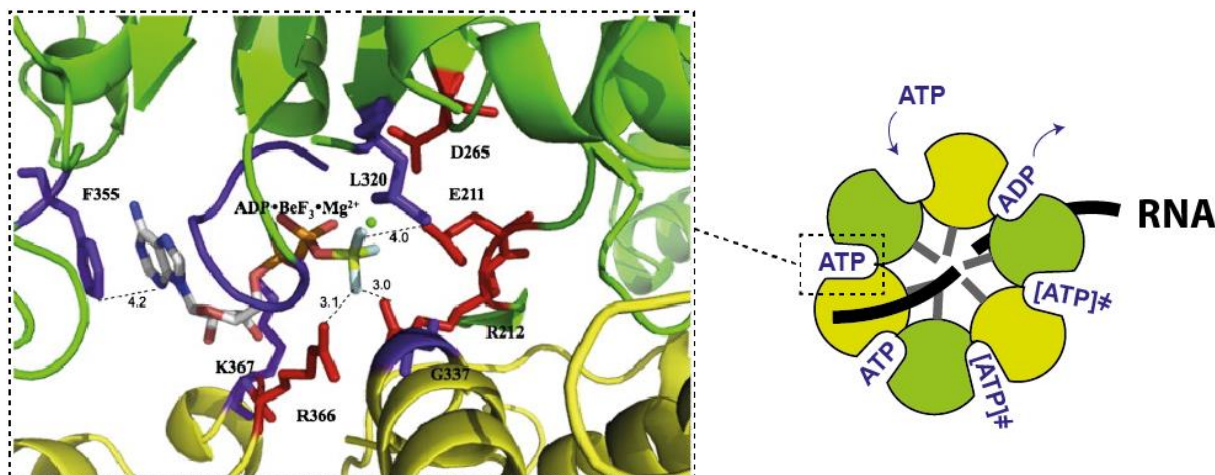


Figure 33 : Poches ATPase formées aux interfaces entre protomères EcRho. La poche ATPase située à l'interface des sous-unités D et E (en vert et jaune) de la structure en anneau asymétrique (PDB: 3ICE) est détaillée ici. Les principaux acides aminés impliqués dans le mécanisme d'hydrolyse sont indiqués. La fonction de ceux représentés en rouge a été validée par mutagenèse dirigée (Balasubramanian & Stitt, 2010). Au centre, l'analogue ADP•BeF₃ est montré avec l'ion Mg²⁺ associé (bille verte) (adapté de Balasubramanian & Stitt, 2010)).

Dans deux des structures cristallines d'EcRho, toutes les poches ATPase sont occupées par un analogue de l'ATP. Dans la structure en anneau ouvert (**Figure 30**, structure de gauche), les poches ATPase sont occupées par des molécules d'AMP-PNP, toutes dans une configuration qui ne paraît pas catalytiquement productive (Skordalakes & Berger, 2003). Cette observation renforce l'idée que l'anneau ouvert représente un état précoce (pré-catalytique) du cycle réactionnel d'EcRho (**Figure 30**) (Skordalakes & Berger, 2003). Dans la structure asymétrique fermée d'EcRho (**Figure 32**), les conformations des poches ATPase (qui contiennent un analogue nucléotidique ADP•BeF₃) ne sont pas toutes à fait identiques, globalement définissant quatre états distincts censés représenter quatre étapes du cycle d'hydrolyse de l'ATP : **T** (triphosphate) lorsque la poche est occupée par une molécule d'ATP (état initial de la réaction); **T***, lorsque la molécule d'ATP approche l'état de transition de la réaction d'hydrolyse; **D** (diphosphate), lorsque la poche contient les produits d'hydrolyse (ADP + Pi); **E** (exchange) lorsque les produits d'hydrolyse sont remplacés par une nouvelle molécule d'ATP (**Figure 32**) (Thomsen & Berger, 2009). Ces observations sont compatibles avec une étude combinant des mesures à l'équilibre et en conditions de cinétique transitoire ("quench flow") qui indique qu'en moyenne cinq molécules de nucléotide cofacteur sont stablement associées à chaque hexamère EcRho présent en solution (Adelman *et al.*, 2006).

2. LES ACTIVITÉS ENZYMATIQUES DE RHO

A. L'activité NTPase

Le facteur EcRho ne possède une activité NTPase qu'en présence d'ARN (Lowery & Richardson, 1977; Lowery-Goldhammer & Richardson, 1974). Les cofacteurs assurant la meilleure activation sont les polymères poly[rC] d'une longueur supérieure ou égale à ~300 nt (Rabhi *et al.*, 2010a). EcRho est capable d'hydrolyser efficacement les quatre rNTP et le dATP (en présence d'ARN) qui, néanmoins, induisent des changements conformationnels sensiblement différents au sein d'EcRho (Brennan *et al.*, 1990). Les conséquences de ces différences sur le rendement mécano-chimique de l'enzyme sont débattues (Brennan *et al.*, 1990; Walmacq *et al.*, 2004). Néanmoins les facteurs Rho d'autres espèces peuvent présenter une spécificité pour le substrat NTP plus marquée (voir page 101).

Le nombre de sites (3 ou 6) au sein d'EcRho capables de fixer et/ou d'hydrolyser l'ATP ainsi que le mécanisme d'hydrolyse (concerté ou séquentiel) ont été longtemps débattus (Browne *et al.*, 2005; Browne & Stitt, 2005; Kim & Patel, 1999; Kim *et al.*, 1999; Stitt, 1988; Stitt, 2001). Actuellement, un consensus semble se dégager autour des caractéristiques de la structure d'EcRho organisée en anneau asymétrique (**Figure 32**) (Thomsen & Berger, 2009) : six poches ATPase (dont cinq stablement occupées par un nucléotide) seraient impliquées à tour de rôle dans un mécanisme d'hydrolyse séquentiel (Adelman *et al.*, 2006; Rabhi *et al.*, 2010b). Le relargage des produits d'hydrolyse – Pi (Adelman *et al.*, 2006) ou ADP (Chen & Stitt, 2009) – par EcRho constituerait l'étape cinétiquement limitante du cycle ATPase. Aucune donnée expérimentale ne permet actuellement d'évaluer précisément ces éléments importants du mécanisme d'hydrolyse pour des facteurs Rho phylo-divergents.

B. L'activité hélicase et importance du site *Rut*

En 1987, Brennan et collaborateurs observent qu'EcRho est capable de provoquer la dissociation d'un hybride ARN:ADN de 28 pb formé à partir d'un transcrit de la fin de l'opéron *trp* d'*E. coli* (qui contient un terminateur Rho-dépendant) (Brennan *et al.*, 1987). Cette activité hélicase, qui dépend strictement de l'hydrolyse de NTP par EcRho, n'est pas limitée aux substrats dérivés de *trp* mais nécessite que le brin ARN contienne un site *Rut*

(Brennan *et al.*, 1987; Brennan *et al.*, 1990; Walmacq *et al.*, 2004; Walstrom *et al.*, 1997a; Walstrom *et al.*, 1997b).

Aujourd'hui, les caractéristiques des sites *Rut* restent mal définies et il n'est pas clair si celles-ci sont conservées d'espèce à espèce. A partir de l'analyse des quelques terminateurs Rho-dépendants bien définis, il a été proposé que, pour EcRho, **un site *Rut* est une séquence ARN d'au moins 77 nt de long, riche en cytosines, pauvre en guanines, et contenant globalement peu de structures secondaires** (Alifano *et al.*, 1991; Schneider *et al.*, 1993). Depuis, des séquences présentant un biais C>G ont été retrouvées en amont des loci de terminaison Rho-dépendante identifiés chez *E. coli* par analyse transcriptomique (Peters *et al.*, 2012). Malheureusement, ces informations restent trop vagues pour définir une séquence *Rut* consensus. Néanmoins, plusieurs équipes ont développé des séquences *Rut* synthétiques pour faciliter l'étude des propriétés d'EcRho (Guerin *et al.*, 1998; Park & Roberts, 2006). Notamment, il a pu être montré qu'une séquence d'environ 50 nt très riche en pyrimidine était capable d'induire la terminaison de la transcription par EcRho *in vitro* et *in vivo* (Guerin *et al.*, 1998). Une séquence similaire de 57 nt, appelée *aRut* (*artificial Rut*) a été utilisée pour créer des substrats ARN:ADN "minimaux" permettant l'étude rigoureuse de l'activité hélicase d'EcRho *in vitro* (Walmacq *et al.*, 2004; Walmacq *et al.*, 2006). Ces substrats ont notamment servi à démontrer que le déplacement ATP-dépendant d'EcRho le long d'un brin ARN n'intervenait que dans la direction 5'→3' (Walmacq *et al.*, 2004). Ils ont permis également de démontrer qu'EcRho dissociait le brin d'ADN par un mécanisme d'exclusion stérique (le canal central de l'hexamère ne pouvant accommoder que le simple brin ARN) et avec une processivité modérée (60-80 pb) ou encore que la formation initiale d'un complexe actif (incluant reconnaissance du site *Rut* par le PBS, entrée de la chaîne ARN dans le canal central, et contact avec le SBS) constituait l'étape cinétiquement limitante de la réaction hélicase (Schwartz *et al.*, 2007a; Walmacq *et al.*, 2004; Walmacq *et al.*, 2006). De plus, l'utilisation de substrats ARN biotinylés contenant un site *aRut* a confirmé qu'EcRho était une ARN translocase directionnelle et démontré que cette dernière était capable de dissocier efficacement des complexes biotin:streptavidine en présence d'ATP, suggérant que l'enzyme opère par un mécanisme non-spécifique de type "buldozer" (Schwartz *et al.*, 2007a).

Des substrats ARN:ADN "tripartites" ont également été construits pour explorer le mécanisme de translocation directionnelle d'EcRho (**Figure 34**).

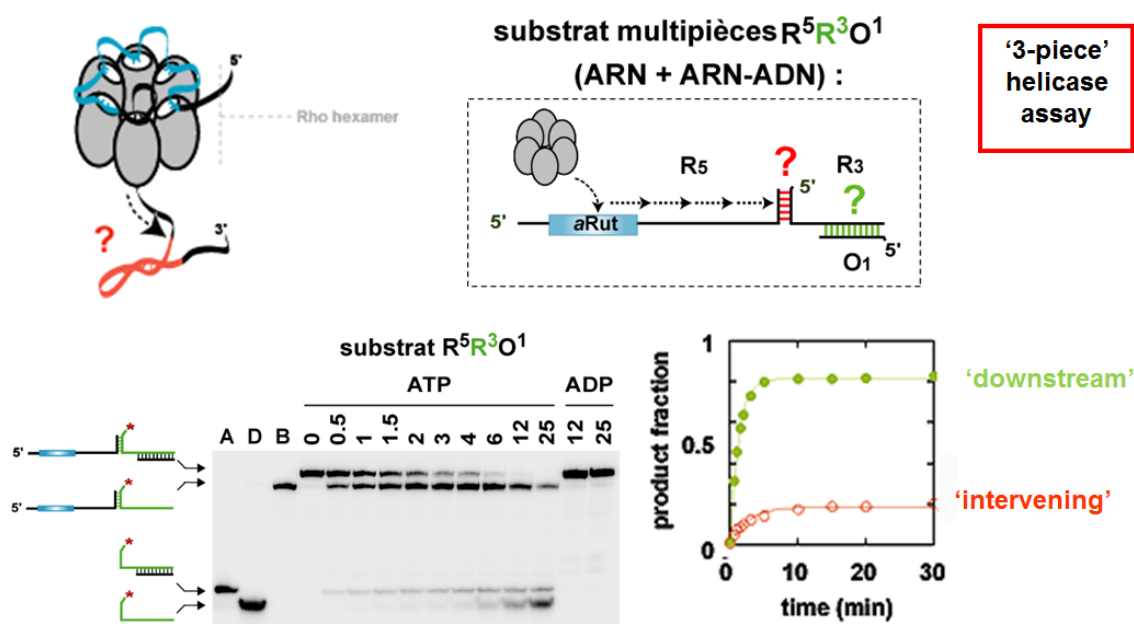


Figure 34: *Activité hélicase d'EcRho avec des substrats ARN:ADN tripartites.* Les substrats sont composés de deux brins d'ARN (R3 et R5) et d'un oligodésoxyribonucléotide (O1). La dissociation de l'oligonucléotide O1 représente la voie réactionnelle majoritaire ("downstream") lorsque les substrats sont incubés en présence d'EcRho et d'ATP. La dissociation de la jonction ARN:ARN ("intervening") est un événement plus rare (d'après Schwartz *et al.*, 2007b).

De façon surprenante, deux voies réactionnelles concurrentes (**Figure 34**) ont été mises en évidence dans ce cas (Schwartz *et al.*, 2007b). Encore plus surprenant, la voie réactionnelle attendue – correspondant à la fixation d'EcRho sur le site *aRut*, à sa translocation vers l'extrémité 3' et à la fusion du premier duplexe rencontré (duplexe ARN:ARN, en rouge sur la **Figure 34**) – est la voie minoritaire (~20%). Cette voie peut être réprimée par la délétion de la séquence *aRut* ou son appariement avec un oligonucléotide complémentaire, ce qui n'est pas le cas pour la voie majoritaire (~80% ; (Schwartz *et al.*, 2007b). En empruntant cette dernière, EcRho semble capable d'éviter la jonction ARN:ARN et de provoquer la dissociation de l'hybride ARN:ADN situé plus en aval (**Figure 34**). Ce résultat inattendu découle de la capacité d'EcRho à accommoder préférentiellement la jonction ARN:ARN (qui ressemble à une structure tige-boucle) sur son PBS lors de son interaction initiale avec le substrat (**Figure 35**) (Schwartz *et al.*, 2007b). Bien que cette observation semble en contradiction avec la définition d'un site *Rut* (pauvre en structures secondaires; voir ci-dessus), certains terminateurs Rho-dépendants comme le terminateur tR1 du phage λ contiennent des motifs en tige-boucle très stables (**Figure 35**) qui peuvent s'avérer fonctionnellement importants (Vieu & Rahmouni, 2004).

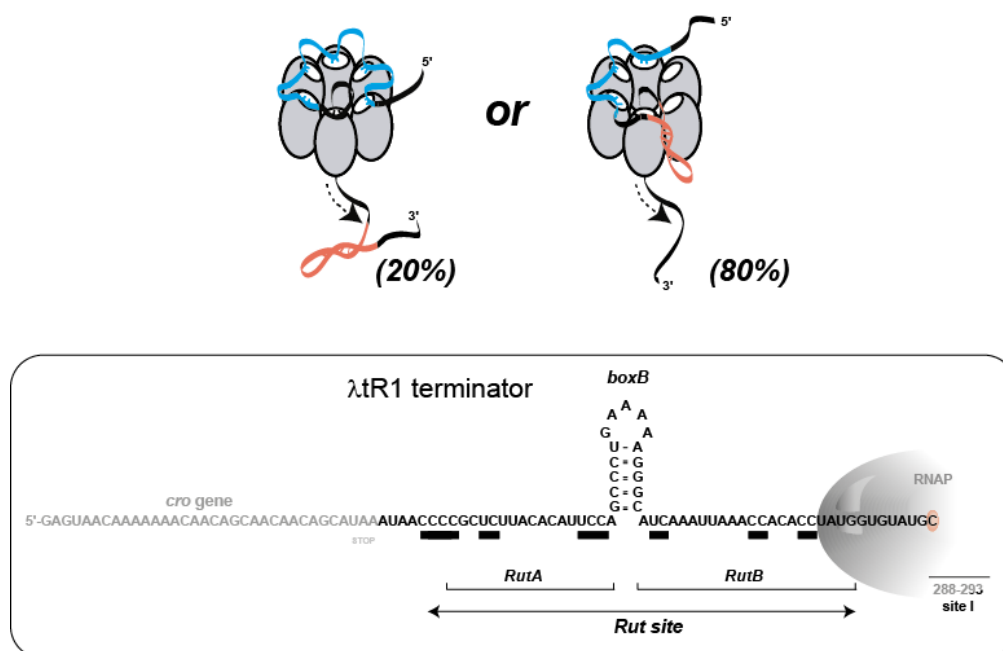


Figure 35 : Topologie des complexes Rho/ARN potentiellement formés lorsqu’une structure secondaire stable est présente dans la chaîne ARN. L’espèce majoritairement formée (en haut à droite) est stabilisée par l’accommodation de la structure ARN (rouge) au sein de son PBS. Les détails moléculaires de cette interaction ne sont pas connus. Le terminateur λ tR1 (situé en aval du gène *cro*) contient une structure en tige-boucle (box B) qui divise son site Rut en deux sous-sites fonctionnels et pourrait être reconnu par Rho suivant cette topologie particulière. Les dinucléotides 5’-YC susceptibles d’être reconnus d’une façon ‘canonique’ par le PBS de Rho sont soulignées en noir (adapté de Schwartz *et al.*, 2007b).

Une recherche de ligands ARN ayant une forte affinité pour EcRho par la méthode SELEX a également généré deux classes de molécules compatibles avec les observations précédentes: *i*) des ARN simple brins riches en C et *ii*) des motifs ARN en tige-boucle (Schneider *et al.*, 1993). Depuis, des terminateurs Rho-dépendants ont été retrouvés associés à des motifs de type “riboswitch” (voir page 92) fortement structurés et dépourvus de sites *Rut* canoniques (Hollands *et al.*, 2012; Takemoto *et al.*, 2015). Ces données démontrent qu’EcRho est capable de reconnaître une chaîne ARN de différentes façons, ce qui complique l’identification des sites *Rut*. Néanmoins, cette propriété, associée à l’activité de translocase/hélicase d’EcRho, pourrait jouer un rôle physiologique important, notamment dans le démantèlement Rho-dépendant et à l’échelle du génome de structures ADN:ARN en “R-Loop” (Gowrishankar *et al.*, 2013; Leela *et al.*, 2013) ou de CET arrêtés (Dutta *et al.*, 2011) à des loci dépourvus de sites *Rut* canoniques.

3. LE MECANISME DE TERMINAISON DE LA TRANSCRIPTION RHO-DEPENDANTE

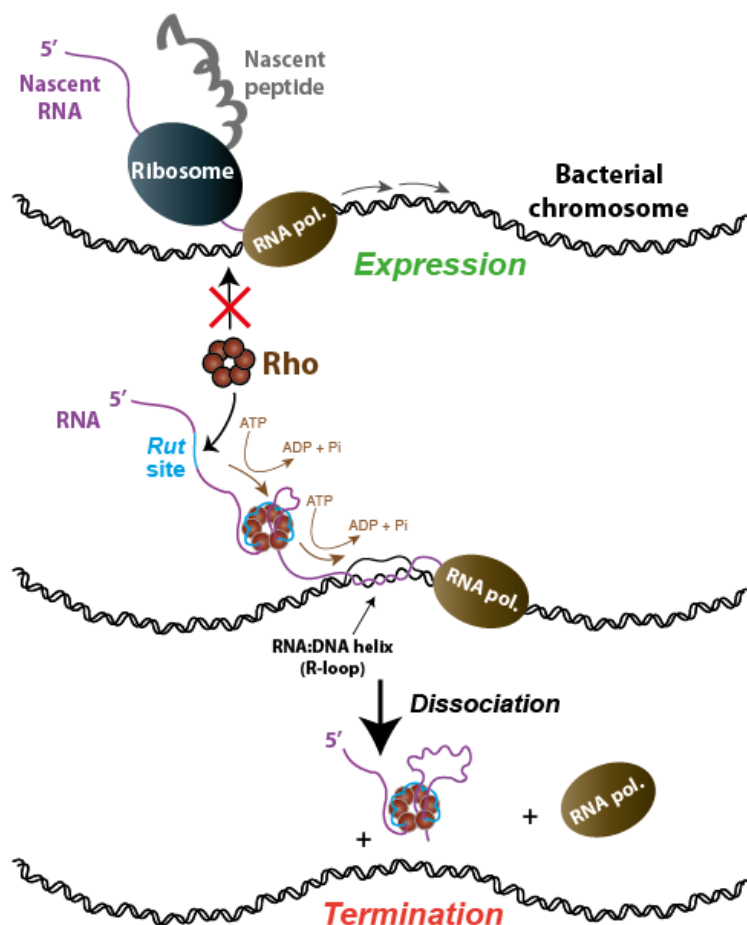


Figure 36 Modèle de terminaison de la transcription Rho-dépendante.

Lorsque le transcrit naissant (ou une région de celui-ci) est nu, Rho va pouvoir s'y fixer au niveau d'un site Rut (en bleu). Cette interaction va activer la fonction ATPase de Rho qui va utiliser l'énergie résultante pour se déplacer le long du transcrit en direction de l'ARNP. Ce mouvement de translocation va permettre à Rho de rattraper l'ARNP et d'induire la dissociation du CET. En cours de translocation, Rho peut également être amené à dissocier des structures de type "R-Loop".

Pour que la terminaison Rho-dépendante puisse avoir lieu, deux conditions préalables doivent être réunies:

i) L'ARN naissant doit contenir une séquence *Rut* pour que Rho puisse s'y fixer via son PBS puis former un complexe catalytiquement compétent via une seconde interaction avec son SBS (voir page 74). Il a été proposé récemment que le facteur NusG stimule la terminaison Rho-dépendante en facilitant ce processus de reconnaissance et d'activation à des sites *Rut* considérés comme non-optimaux (Peters *et al.*, 2012; Shashni *et al.*, 2014). Les modèles plus anciens proposaient, au contraire, que NusG stimule la dernière étape du processus de terminaison (dissociation de l'ARNP - voir plus bas) (Burns *et al.*, 1999; Dutta *et al.*, 2008).

ii) La région du transcrit naissant contenant le site *Rut* doit impérativement être accessible et notamment dépourvue de ribosome (**Figure 36**). En plus de masquer le site *Rut*,

le ribosome (via la protéine NusE/S10) séquestre le CTD de NusG (voir Chapitre I page 41), soustrayant ainsi ce facteur à son action stimulante de “pontage” entre Rho et l’ARNP (voir Chapitre I page 32) (Peters *et al.*, 2011). Rho semble jouer un rôle important dans la terminaison de la transcription des ARN non-traduits (Peters *et al.*, 2009) même si la transcription de certains d’entre-deux, comme les ARNr, est protégée par un complexe d’anti-terminaison (voir Chapitre I page 34). La transcription de certaines régions 5’UTR peut également faire l’objet de phénomènes d’atténuation de la transcription Rho-dépendante (Bossi *et al.*, 2012; Hollands *et al.*, 2012) tout comme la transcription de régions intergéniques (Peters *et al.*, 2009; Peters *et al.*, 2012). Enfin, l’absence de traduction explique certainement pourquoi Rho réprime préférentiellement la transcription antisense (Peters *et al.*, 2009; Peters *et al.*, 2012). La mise à nu d’un site *Rut* peut parfois résulter d’une rupture du couplage traduction/transcription due à la présence d’une mutation non-sens (Richardson *et al.*, 1975), à l’insertion d’ADN étranger, à une carence en acide(s) aminé(s) (Gong & Yanofsky, 2003), ou à l’action d’un petit ARN non-codant (Bossi *et al.*, 2012). Dans ces cas, la terminaison Rho-dépendante interviendra au sein de l’opéron, diminuant ainsi l’expression des gènes situés en aval. Ce phénomène est appelé “polarité transcriptionnelle”.

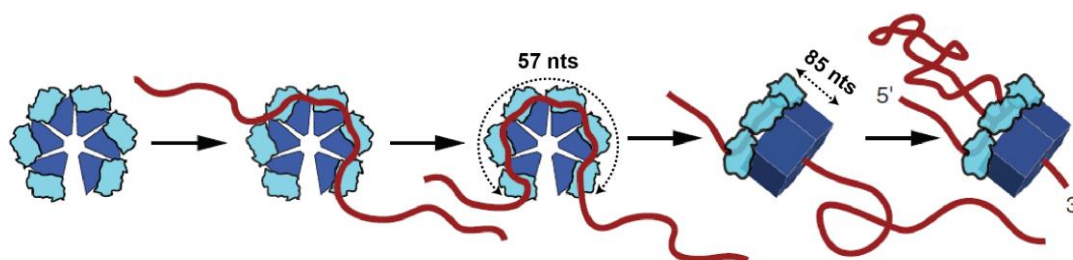


Figure 37 : Interaction entre EcRho et l’ARN. La fixation du site *Rut* sur EcRho (jusqu’à 6 sous-unités) provoque l’entrée du brin ARN au centre de l’hexamère, où il vient contacter le SBS. EcRho adopte alors une conformation active (fermeture de l’anneau, hydrolyse de trois molécules d’ATP pour adopter la configuration asymétrique), avant de se mouvoir le long du transcrit d’une manière ATP-dépendante tout en maintenant son interaction avec le site *Rut* (tethered tracking). Les flèches et les chiffres correspondent aux empreintes ARN observées par nano-manipulation (d’après Koslover *et al.*, 2012).

Les différentes étapes constituant le processus de terminaison Rho-dépendante ont été étudiées en détail pour EcRho (**Figure 36**). La formation d’un complexe actif entre EcRho et son substrat ARN est un processus lent (Schwartz *et al.*, 2007b; Walmacq *et al.*, 2006) et multi-étapes (Kim & Patel, 2001). Une fois fixé au PBS de l’enzyme via un site *Rut*, l’ARN doit pénétrer au centre de l’anneau hexamérique pour y contacter le SBS (**Figure 30 & Figure 37**). Puis, EcRho doit adopter une conformation active “fermée”, certainement ressemblant à la structure “asymétrique” (**Figure 32**) (Soares *et al.*, 2014; Thomsen & Berger, 2009). L’acquisition de cette conformation nécessiterait l’hydrolyse de trois molécules d’ATP sans

translocation de la part de l'enzyme (Kim & Patel, 2001; Soares *et al.*, 2014). C'est seulement après cette dernière étape d'activation qu'EcRho serait capable de transloquer l'ARN en son sein, nucléotide par nucléotide, suivant le principe d'escorte ATP-dépendante (**Figure 32.C**) décrit plus haut (page 76) (Enemark & Joshua-Tor, 2006; Thomsen & Berger, 2009)

Deux études récentes de nano-manipulation d'EcRho à l'échelle de la molécule unique ont observé que la translocation de Rho s'accompagne d'une diminution de la distance entre les deux extrémités de la chaîne ARN, en accord avec un mécanisme de translocation par “*tethered tracking*” dans lequel le site Rut reste fixé au PBS (**Figure 37**) (Gocheva *et al.*, 2015; Koslover *et al.*, 2012). Ce mécanisme implique qu'il existe une asymétrie au sein de l'hexamère au cours de la translocation car seules cinq des six interfaces peuvent être stabilisées par l'ARN joignant les sites PBS de sous-unités adjacentes (**Figure 37**). Ainsi, il existe au moins une interface au sein de l'hexamère qui est intrinsèquement moins stable (Soares *et al.*, 2014). En cours de translocation, cette interface “faible” est particulièrement vulnérable lorsqu'elle n'est pas stabilisée par un nucléotide cofacteur (**Figure 32.B & Figure 38**, état “*exchange*” **E**). C'est dans cet état précis que l'une des deux sous-unités de l'interface doit réengager un nouveau nucléotide de la chaîne ARN en contactant tout d'abord son groupement 2'-OH (Thomsen & Berger, 2009).

Ces caractéristiques, conjuguant translocation par “*RNA escort*” et “*tethered tracking*”, expliqueraient pourquoi EcRho est particulièrement sensible à l'absence d'un groupement 2'-OH tous les ~7 nt (**Figure 38**) (Schwartz *et al.*, 2009; Soares *et al.*, 2014). La vulnérabilité périodique de l'interface “faible” pourrait constituer un élément de régulation du cycle de translocation, où la difficulté à engager un nouveau nucléotide (via son groupement 2'-OH) pourrait plus facilement conduire à l'ouverture de l'anneau hexamérique et à la dissociation du complexe EcRho:ARN (**Figure 38**) (Soares *et al.*, 2014). Cette difficulté pourrait être modulée par l'identité des résidus ARN occupant le canal central de l'hexamère dont l'exiguïté devrait rendre le processus de translocation particulièrement sensible aux contraintes stériques (Soares *et al.*, 2014).

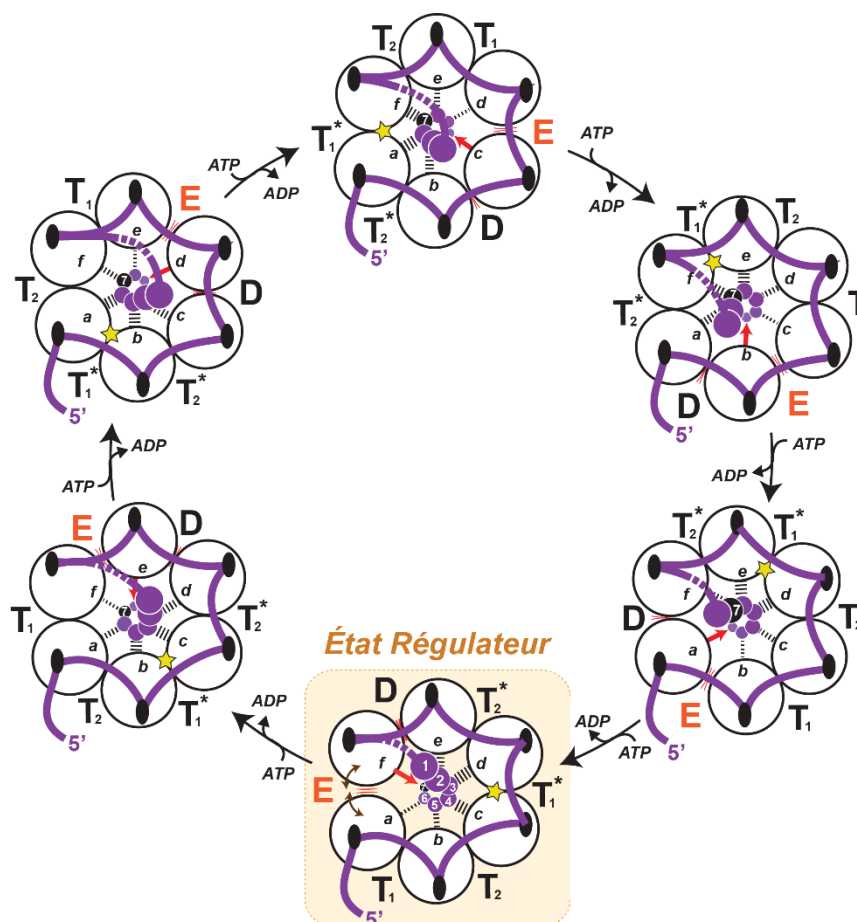


Figure 38 : Modèle du cycle de translocation pour EcRho. La circulation de la chaîne ARN (en violet) sur le PBS (constitué d'un site par sous-unité, représenté par un ovale noir) et le mécanisme de "tethered tracking" induisent une asymétrie dans l'hexamère qui contient une interface "faible" (dont les sous-unités ne sont pas reliées par un segment d'ARN). Au cours du cycle de translocation, cette interface "faible" peut plus facilement s'ouvrir lorsqu'elle n'est pas stabilisée par un nucléotide cofacteur (état **E**), définissant ainsi un état régulateur pouvant conduire à l'inactivation de l'enzyme et la dissociation de l'ARN. Les lignes pointillées noires symbolisent les contacts entre le SBS et les nucléotides de la chaîne ARN représentés par des ronds violets. Les flèches rouges représentent les premiers contacts formés entre EcRho et chaque nouveau nucléotide (impliquant son groupement 2'-OH) (d'après Soares et al., 2014).

La vitesse de translocation de Rho est un paramètre important car, selon le modèle de compétition cinétique, la terminaison Rho-dépendante dépend de la capacité de Rho à rattraper l'ARNP (Jin *et al.*, 1992). Cette compétition cinétique est corroborée par l'effet que la vitesse de translocation de l'ARNP (qui peut être modulée par des mutations) a sur l'efficacité de terminaison Rho-dépendante. Par exemple, un mutant ARNP possédant une vitesse d'élongation accrue (réduite) exhibe un profil de terminaison "sauvage" uniquement en présence d'un mutant Rho dont l'activité enzymatique est également accrue (réduite) (Guarente & Beckwith, 1978; Jin *et al.*, 1988; Jin *et al.*, 1992). Récemment, la vitesse de translocation d'EcRho a été évaluée à ~ 60 nt/s dans des conditions de nano-manipulation où l'ARNP progresse 2 à 6 fois moins vite (Gocheva *et al.*, 2015). Ces données sont donc compatibles avec le modèle de compétition cinétique (Jin *et al.*, 1992) et semblent adéquates

pour que l'ARNP soit rattrapée avant qu'EcRho ne se dissocie de l'ARN, soit après 60-80 nt transloqués (Gocheva *et al.*, 2015; Walmacq *et al.*, 2004). Néanmoins, la dissociation du CET par EcRho semble intervenir préférentiellement à des positions où l'ARNP marque une pause (Galloway & Platt, 1988; Hollands *et al.*, 2014; Kassavetis & Chamberlin, 1981; Morgan *et al.*, 1983; Stewart *et al.*, 1986). Il n'est pas clair si cela reflète simplement la compétition cinétique entre EcRho et l'ARNP ou bien indique que la conformation adoptée par l'ARNP à un site de pause est plus sensible à l'action d'EcRho.

La dissociation de CET par le facteur Rho est aujourd'hui encore l'étape du mécanisme de terminaison de la transcription suscitant le plus de débat. Trois modèles principaux, ressemblant fortement aux modèles de terminaison intrinsèque (voir page 28), ont été proposés :

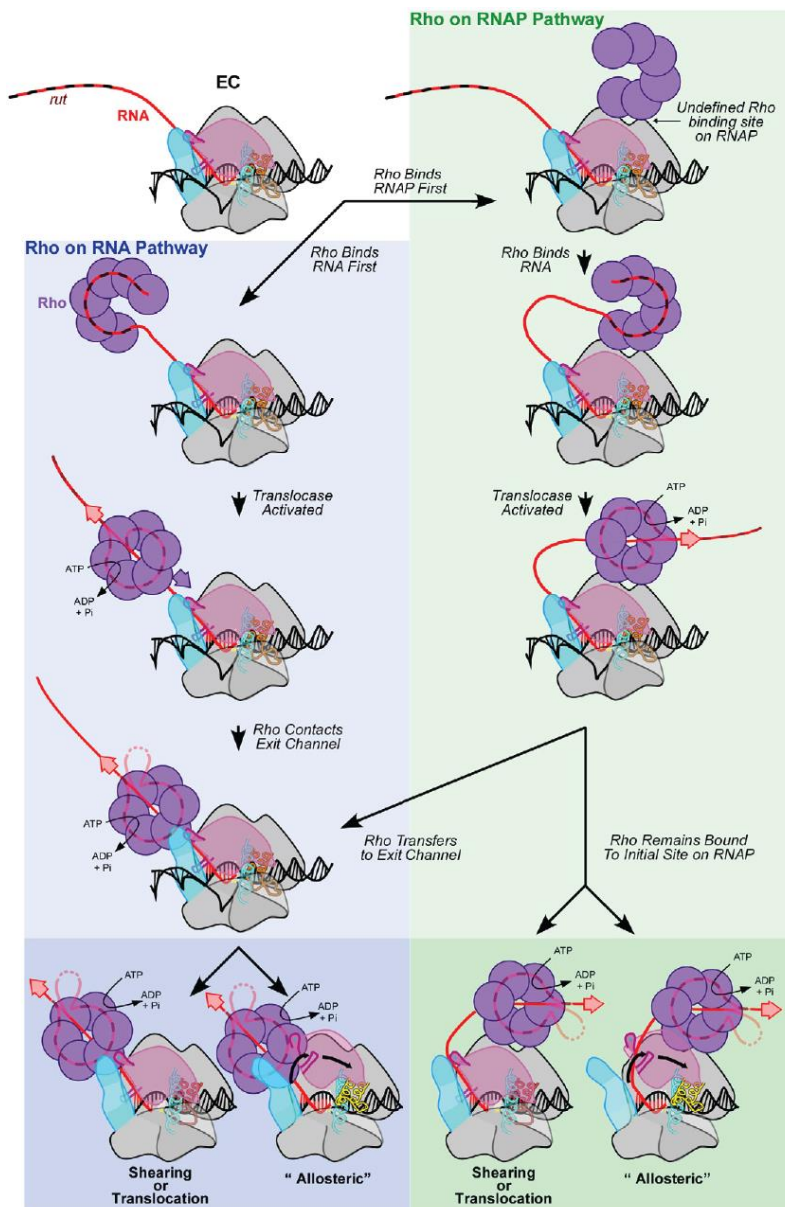


Figure 39 : Modèles de dissociation du CET par EcRho. A gauche, est représentée l'hypothèse selon laquelle EcRho se lie seulement à l'ARN (fond bleu). A droite, l'hypothèse alternative stipule qu'EcRho est en interaction, permanente avec l'ARNP (fond vert) En bas, sont représentés les différents modèles de dissociation du CET décrits dans le texte (d'après Peters *et al.*, 2011).

- Le modèle par “*hybrid shearing*” (**Figure 39**) stipule que, une fois en contact avec l’ARNP, EcRho va tirer sur le brin ARN jusqu’à provoquer la dissociation de l’hybride ARN:ADN au cœur de l’ARNP (Richardson, 2002).
- Le modèle par “*(hyper)translocation*” (**Figure 39**) propose que l’ARNP est poussée en avant le long de l’ADN par la translocation d’EcRho en appui sur elle. Ce mouvement, effectué sans ajout concomitant de NTP à l’extrémité 3’ du transcrit, est censé provoquer le raccourcissement progressif de l’hybride ARN:ADN et l’effondrement de la bulle de transcription conduisant à la dissociation du CET (Park & Roberts, 2006). Ce modèle est notamment supporté par le fait qu’un CET rétro-transloqué est plus résistant à l’action d’EcRho (Dutta *et al.*, 2008).
- Le modèle allostérique (**Figure 39**) propose que l’action d’EcRho provoque l’inactivation de l’ARNP puis de nombreux réarrangements allostériques conduisant à l’ouverture de la pince β/β' et à la perte de la plupart des contacts de l’ARNP avec l’ARN et l’ADN (Dutta *et al.*, 2011).

Aucun de ces modèles ne fait aujourd’hui consensus (Peters *et al.*, 2011), la situation étant compliquée par la proposition récente que Rho et l’ARNP pourraient être en association permanente durant le cycle transcriptionnel (**Figure 39**). En effet, il est rapporté qu’EcRho peut interagir avec des CET fixés sur billes magnétiques et contenant des transcrits ARN trop courts pour émerger du canal secondaire de l’ARNP (et donc pour former une interaction directe avec Rho) (Dutta *et al.*, 2011). Dans la même étude, EcRho ne peut induire la terminaison de la transcription avec des CET préalablement incubés avec des mutants EcRho inactifs, signifiant que ces mutants restent fixés à l’ARNP et empêchent l’interaction avec EcRho sauvage (Epshtein *et al.*, 2010). La possibilité qu’EcRho soit un composant permanent du CET est supportée par des données d’immunoprécipitation à l’échelle du génome (“*ChIp on chip*”) démontrant une co-localisation de l’ARNP et d’EcRho le long des unités transcriptionnelles d’*E. coli* (Mooney *et al.*, 2009a). Cependant, une étude *in vitro* récente conteste ce modèle d’interaction permanente, les auteurs observant une interaction stable entre EcRho et l’ARNP uniquement quand le CET contient un ARN assez long (123 nt) et pourvu d’un site *Rut* (Kalyani *et al.*, 2011). Des expériences de nano-manipulation à l’échelle de la molécule unique n’ont pas non plus mis en évidence d’interaction permanente entre EcRho et l’ARNP (Koslover *et al.*, 2012) dont l’existence reste, donc, encore sujet à débat.

4. LES FONCTIONS DU FACTEUR RHO

Au-delà de sa fonction consistant à délimiter de manière constitutive l'extrémité 3' de certaines unités transcriptionnelles (voir Chapitre I page 30), Rho est également impliqué dans d'autres fonctions majeures chez *E. coli* et probablement chez les espèces apparentées comme *S. enterica*.

A. Régulation (conditionnelle) de la transcription

Il est depuis longtemps établi que Rho est capable de réguler l'expression de certains gènes, via le phénomène de polarité transcriptionnelle (Richardson *et al.*, 1975). Un exemple original de régulation par polarité transcriptionnelle concerne l'opéron *tnaCAB* d' *E. coli* dont l'expression est modulée par la concentration intra-cytoplasmique de tryptophane. Cet opéron, codant pour des protéines impliquées dans le catabolisme du tryptophane, comprend également dans sa partie amont la séquence codant un peptide leader, TnaC, dont le codon stop se situe juste avant un site *rut*. La présence de tryptophane en forte concentration inhibe les facteurs de relargage RF1 et RF2 qui, habituellement, clivent le peptide TnaC et permettent la dissociation du ribosome. Cette inhibition provoque la fixation prolongée du ribosome sur le codon stop, recouvrant ainsi le site *Rut* et inhibant l'action de Rho. En condition de carence en tryptophane, le peptide *tnaC* est correctement clivé, ce qui entraîne la dissociation du ribosome ainsi que la libération du site *Rut* pour permettre l'arrêt prématuré de la transcription (Gong & Yanofsky, 2003).

Un exemple similaire concerne le gène *rho* lui-même. Chez *E. coli*, *rho* est précédé d'une longue région 5'-UTR (255 bp) contenant une séquence de 102 bp codant pour un petit peptide leader (34 aa) dénommé RhoL. La séquence *rhoL* contient en son sein plusieurs sites de terminaison qu'EcRho utilise pour réguler l'expression de son propre gène *rho* (Barik *et al.*, 1985; Brown *et al.*, 1982; Matsumoto *et al.*, 1986). Cette autorégulation permet à EcRho d'être présent à un taux relativement constant (~ 1000 hexamères/cellule) tout au long du cycle cellulaire (Blumenthal *et al.*, 1976; Geiselmann *et al.*, 1992). Bien que ces caractéristiques suggèrent un mécanisme d'atténuation reposant sur un phénomène de polarité transcriptionnelle, il n'a pas encore été formellement démontré que la traduction de *rhoL* est requise pour l'autorégulation d'EcRho. Un mécanisme similaire d'autorégulation de la concentration en Rho a été mis en évidence chez *B. subtilis* (Ingham *et al.*, 1999).

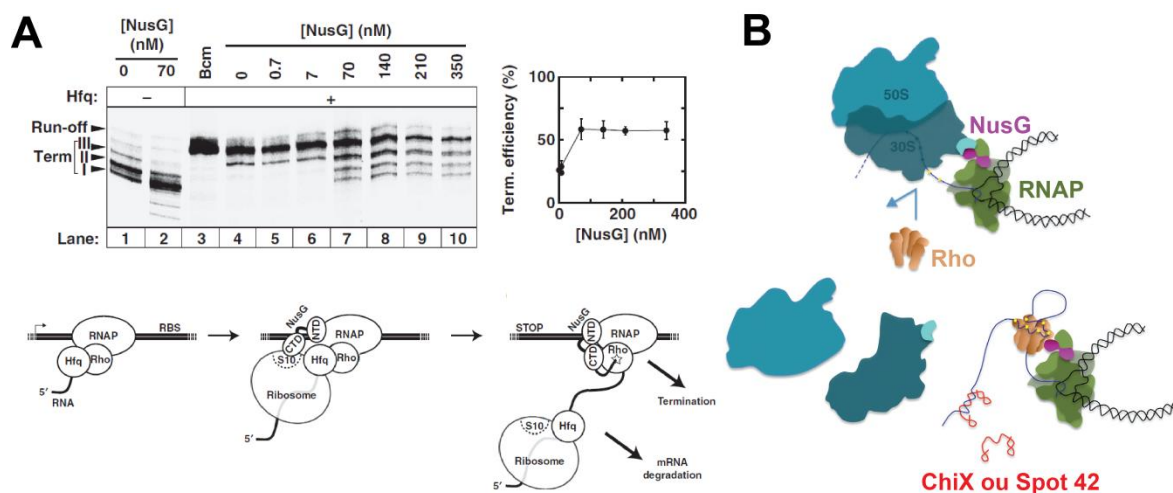


Figure 40 : Implication de Rho dans la régulation sRNA-dépendante. (A) Hfq interagit physiquement avec et inhibe Rho à certains terminateurs Rho-dépendants (ici le terminateur $\lambda tr1$). Cette inhibition, contrecarée par NusG, pourrait contribuer à moduler la fonction de Rho (d'après (Rabhi et al., 2011a)). (B) Durant la transcription d'une région codante, la présence d'un ribosome lié à l'ARNP protège le site Rut (billes jaunes) et le CTD de NusG de l'action de Rho. Lorsqu'un sRNA associé à l'ARNm au niveau du site rbs empêche la fixation des ribosomes, Rho peut alors interagir avec le site Rut et NusG pour assurer la terminaison prématurée de la transcription (adapté de Boudvillain et al., 2013).

Deux études récentes indiquent que la polarité transcriptionnelle peut également intervenir dans des mécanismes de régulation employant des petits ARN non-codants. L'implication de Rho dans ce type de mécanisme était suspectée depuis la mise en évidence que la protéine Hfq (voir Chapitre I page 48) est capable d'inhiber la terminaison Rho-dépendante *in vitro* et *in vivo* (Rabhi et al., 2011a). Cette inhibition est due à une interaction relativement stable entre Hfq et Rho ($K_d \sim 40$ nM) et mutuellement exclusive avec l'interaction Rho:NusG (**Figure 40.A**) (Rabhi et al., 2011a). Les auteurs proposent que l'inhibition de la terminaison Rho-dépendante par Hfq pourrait intervenir durant la transcription des régions 5'-UTR (où NusG n'est pas présent) ou lorsque le couplage traduction/transcription est effectif (**Figure 40.A**). De plus, cette inhibition pourrait être levée par l'action d'un sRNA (Rabhi et al., 2011a), une hypothèse qui n'a cependant pas encore été validée expérimentalement. Néanmoins, l'étude de l'opéron ChiPQ de *S. enterica* a révélé qu'un sRNA, ChiX, était capable *in vivo* d'activer la terminaison Rho-dépendante (Bossi et al., 2012). ChiX se fixe de manière Hfq-dépendante à la région 5'-leader de l'ARNm ChiPQ, à proximité du site *rbs* (Figuerola-Bossi et al., 2009). Ce faisant, ChiX interfère avec la fixation des ribosomes et provoque un phénomène classique de polarité transcriptionnelle via la mise à nu d'un terminateur Rho-dépendant intra-génique situé dans *chiP* (**Figure 40.B**) (Bossi et al., 2012). Cette étude représente la toute première observation de l'implication de Rho dans la régulation sRNA-dépendante. Un mécanisme similaire a été mis en évidence très récemment chez *E. coli* mais, cette fois, le site de fixation du sRNA (Spot 42) est situé dans

une très courte région inter-génique située entre les gènes *galT* et *galK* de l'opéron *galETKM* (Wang *et al.*, 2015). De plus, il semble que dans ce cas le site de fixation du sRNA recouvre partiellement le site *Rut*. Les auteurs proposent que le complexe Spot42:Hfq stimule la dissociation du ribosome à l'extrémité de *galT* (qui autrement pourrait être directement réutilisé pour la traduction de *galK*) et est lui-même relargué durant le processus, ce qui libère le site *Rut* et permet l'atténuation Rho-dépendante de l'expression de *galK* et *galM* (Wang *et al.*, 2015). Comment le complexe Spot42:Hfq puis Rho accèdent au mRNA dans cet environnement "stériquement" très encombré reste néanmoins peu clair.

La polarité transcriptionnelle Rho-dépendante est également impliquée dans des cas où la transcription est potentiellement contre-productive pour la bactérie (limitations des ressources, apparition de mutations délétères, etc.). Par exemple, l'arrêt du ribosome au niveau d'une mutation non-sens provoque le découplage de la traduction et de la transcription et permet à Rho d'agir en aval de la mutation pour limiter le gaspillage des ressources dans la transcription du gène (ou opéron) défectueux (Boudvillain *et al.*, 2013; Richardson, 1991). Ce mécanisme classique d'atténuation Rho-dépendante induite par des mutation non-sens a été observé pour la première fois pour le gène *rho* lui-même (Richardson *et al.*, 1975). Il est par ailleurs intéressant de noter qu'une carence en acides aminés induit la pause du ribosome mais aussi, via la réponse stringente (voire Chapitre I page 40), la pause de l'ARNP à certains loci (Dennis & Nomura, 1974). Pourtant, aucune relation directe entre réponse stringente et activité de Rho n'a encore été décrite.

Plus récemment, deux études ont montré qu'un terminateur Rho-dépendant pouvait constituer le module de régulation de motifs "riboswitch" naturels (Hollands *et al.*, 2012; Takemoto *et al.*, 2015). La réorganisation structurale du motif "riboswitch" induite par l'inducteur (riboflavine ou faible concentration en Mg^{2+} ; **Figure 41**) "libère" un site *Rut* que Rho va pouvoir utiliser pour induire la terminaison de la transcription (Hollands *et al.*, 2012; Takemoto *et al.*, 2015). Il est intéressant de noter que dans les rares cas identifiés, le site *Rut* n'est pas canonique (**Figure 41**), suggérant que Rho peut s'associer productivement à des ARN structurés (voir aussi page 83). Des motifs "riboswitch" Rho-dépendants ont été retrouvés aussi bien chez les γ -Protéobactéries (*E. coli* et *S. enterica*) (Hollands *et al.*, 2012) que chez les Actinobactéries (*Corynebacterium glutamicum*) (Takemoto *et al.*, 2015), des phyla pour lesquels la régulation par riboswitch était prédite pour affecter principalement la traduction (Barrick & Breaker, 2007).

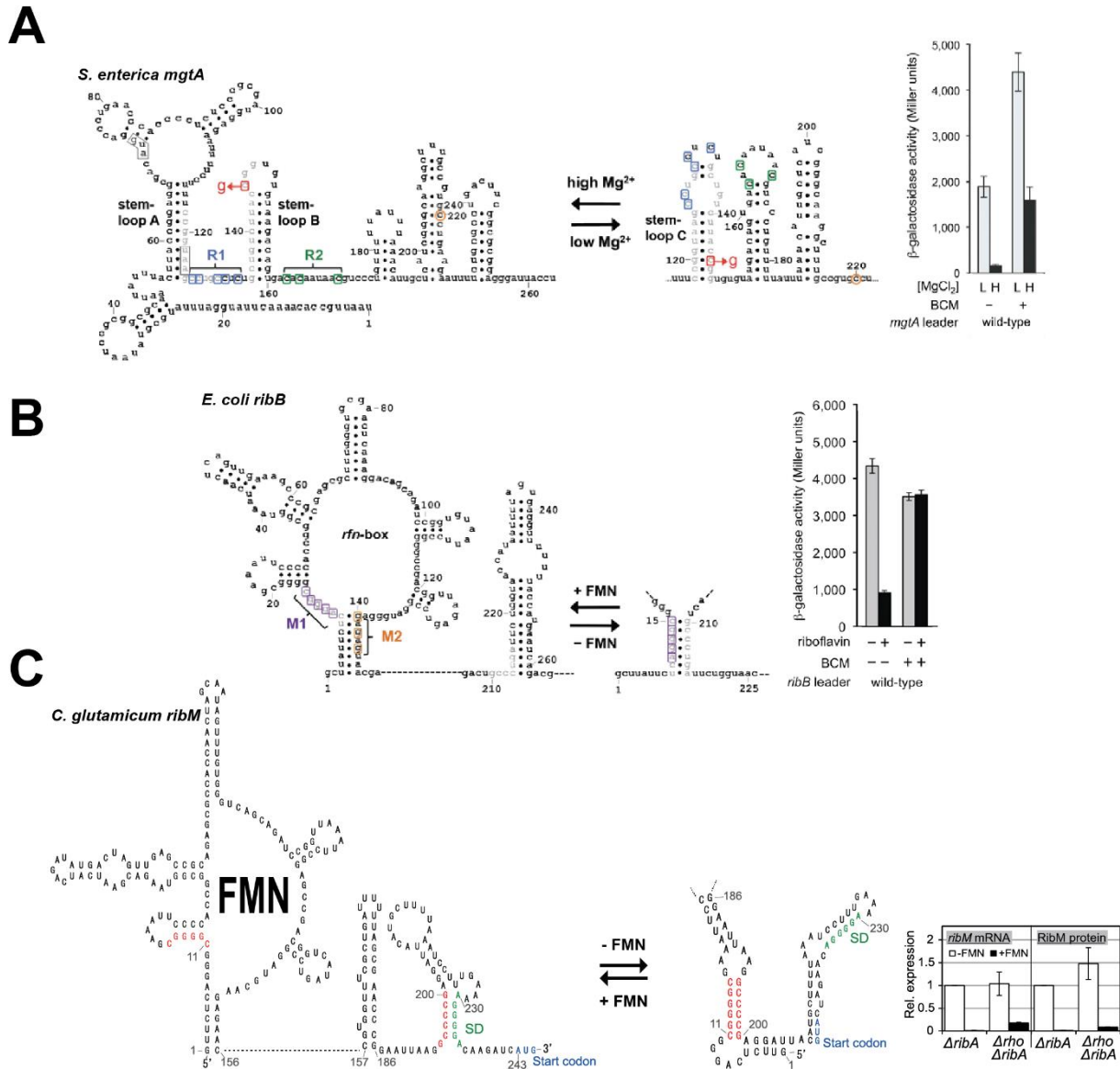


Figure 41 : Motifs “riboswitch” impliquant la terminaison Rho-dépendante. Trois exemples ont été identifiés à ce jour, tous situés dans la région 5'-leader d'un ARNm : du gène *mgtA* chez *S. enterica* (A), du gène *ribB* chez *E. coli* (B), et du gène *ribM* chez *C. glutamicum* (C). Pour chacun, les conformations alternatives de la région 5'-leader sont montrées. Chez *S. enterica* les régions R1 et R2 sont proposées constituer le site Rut. Les diagrammes à droite représentent l'expression génique en fonction de l'inactivation de Rho (BCM, ou Δ *rho*) et de la présence du ligand (Mg^{2+} ou FMN) (adapté de Hollands et al., 2012; Takemoto et al., 2015).

B. Contrôle de la stabilité du transcriptome

Il est maintenant admis que, pour la plupart des organismes vivants, la quasi-totalité du génome est transcrite ou est susceptible de l'être (Libri, 2015). Ce phénomène est attribué en grande partie à la faible spécificité de l'initiation de la transcription qui peut intervenir à de nombreux promoteurs “cryptiques” (Libri, 2015). Cette transcription “anarchique” offre des possibilités de régulation et des avantages évolutifs qui sont intensivement étudiés (Duval et al., 2015; Libri, 2015; Pelechano & Steinmetz, 2013) mais elle peut également avoir des

conséquences délétères. Par exemple, la transcription antisense anarchique de régions contenant un promoteur “authentique” peut inhiber l’expression du gène (ou opéron) contrôlé par ce promoteur, donnant naissance au phénomène dit d’interférence transcriptionnelle (Libri, 2015; Pelechano & Steinmetz, 2013). L’avènement des techniques d’analyse à l’échelle génomique (*ChIp on chip*, RNA-seq) a permis d’impliquer Rho dans le contrôle global de ce phénomène. Il apparaît en effet qu’une fonction majeure de Rho chez *E. coli* et *B. subtilis* est de réprimer la transcription antisense (Nicolas *et al.*, 2012; Peters *et al.*, 2009; Peters *et al.*, 2012). La comparaison des transcriptomes d’*E. coli* cultivés en présence ou non de l’inhibiteur BCM a révélé, par exemple, que 88 % des événements de terminaison Rho-dépendante correspondent à l’inhibition de la transcription antisense (**Figure 11**) (Peters *et al.*, 2012). Cette inhibition est stimulée par la présence de NusG, qui améliorerait la reconnaissance de sites *Rut* déficients, et par la présence de H-NS qui favoriserait la pause du CET (Kotlajich *et al.*, 2015; Landick *et al.*, 2015; Peters *et al.*, 2012).

Plusieurs études suggèrent également qu’EcRho pourrait contribuer à sculpter le transcriptome d’*E. coli* en réprimant largement l’expression de gènes issus de transferts horizontaux (Cardinale *et al.*, 2008; Menouni *et al.*, 2013; Peters *et al.*, 2009). Cette action exercée pour des gènes nocifs, comme le gène *kil* du prophage *rac*, pourrait expliquer pourquoi Rho est essentiel chez *E. coli* (Cardinale *et al.*, 2008), une hypothèse qui est néanmoins contestée (Leela *et al.*, 2013). Un autre point de débat est la raison pour laquelle EcRho ciblerait préférentiellement les prophages (et autres régions acquises par transfert horizontal). Plusieurs hypothèses ont été émises liées *i)* au biais de séquences des prophages (généralement riches en T/A, ce qui favorise leur reconnaissance par H-NS et pourrait être une source de terminateurs Rho-dépendants spécifiques), *ii)* aux sites préférentiels d’intégration (généralement en aval de gènes, notamment des ARNt, provoquant souvent la perte ou inactivation de leurs terminateurs intrinsèques), ou *iii)* au fait que les gènes acquis par transfert horizontal seraient moins efficacement traduits (Cardinale *et al.*, 2008; Landick *et al.*, 2015; Peters *et al.*, 2009).

C. Protection de l’intégrité génomique

Un rôle émergent (et important) de Rho est sa contribution à la protection de l’intégrité génomique qui pourrait résulter de son implication dans la résolution des R-Loops et/ou dans l’élimination des CET arrêtés (pour revue : Gowrishankar *et al.*, 2013). Les R-Loops sont des

hétéroduplexes formés par l'association d'une région d'un transcrit ARN avec la région du brin ADN qui lui est complémentaire. Ces hétéroduplexes pourraient être utilisés comme fragments d'Okazaki par l'ADN polymérase et permettre la réplication de l'ADN (voir Chapitre I page 24) (Kogoma, 1997). La génération de R-Loops est continue chez *E. coli* et doit être régulée (et/ou contrecarrée) au risque d'être létale pour la bactérie (Kogoma, 1997). Les mécanismes de formation des R-Loops sont encore mal définis. Trois modèles principaux ont été proposés (**Figure 42**) :

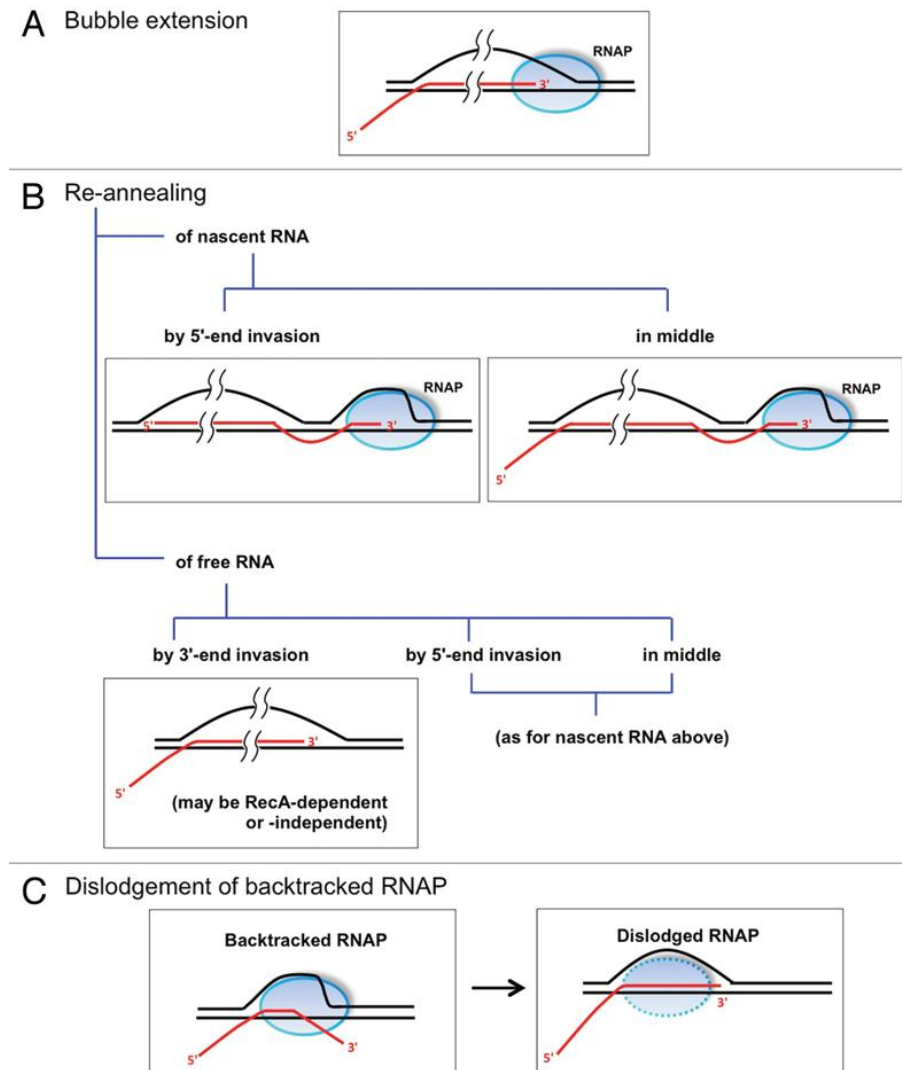


Figure 42 : Modèles de formation des R-Loops.
(d'après Gowrishankar et al., 2013)

- Le modèle par “*bubble extension*” implique que l'ARN néo-synthétisé reste associé avec son brin complémentaire durant l'élongation de la transcription (**Figure 42.A**). Ce modèle ancien, basé sur la formation de R-Loops durant la réplication du plasmide ColE1 (Itoh & Tomizawa, 1980), paraît incompatible avec l'organisation structurale du CET telle qu'on la connaît aujourd'hui, notamment avec la notion de bulle de transcription et l'existence de canaux d'extrusion distincts pour l'ARN et l'ADN (voir **Figure 2**).

- Le modèle par “*re-annealing*” propose que les R-Loops sont dues à la ré-hybridation du brin ADN matrice au transcrit ARN après l’extrusion de ce dernier par le canal secondaire de l’ARNP (**Figure 42.B**). L’envahissement par le transcrit ARN serait favorisé par le surenroulement négatif de l’ADN généré en amont de l’ARNP lors de l’élongation de la transcription (Rahmouni & Wells, 1992). Ceci expliquerait pourquoi la formation de R-loops est favorisée dans des mutants *E. coli* déficients en topoisomerase-I (Masse & Drolet, 1999a; Masse & Drolet, 1999b).
- Le troisième modèle, proposé récemment, stipule qu’une R-Loop pourrait être formée à la suite de la collision entre le complexe de réplication et un CET rétro-transloqué lorsque transcription et réplication ont lieu de façon co-directionnelle (Dutta *et al.*, 2011). Dans ce cas, c’est l’extrémité 3’ du transcrit qui serait impliquée dans l’hétéroduplexe (**Figure 42.C**) mais le chemin réactionnel exact conduisant à la formation de la R-Loop n’est pas précisé (Dutta *et al.*, 2011; Nudler, 2012). Il n’est pas clair si l’extrémité 3’ du transcrit est extrudée (et disponible) via le canal d’entrée des NTP (comme pour un CET rétro-transloqué “classiquement” ; voir page 24) ou bien est appariée au brin ADN matrice au sein même du CET, comme cela a été proposé pour l’initiation de la réplication du bactériophage M13 (Zenkin *et al.*, 2006). Le modèle postule également que le complexe de réplication déplace l’ARNP (**Figure 42.C**) et utilise la R-Loop nouvellement formée pour réamorcer la réplication. C’est ce réamorçage qui expliquerait la présence de cassures dans l’ADN et une instabilité génomique accrue (Dutta *et al.*, 2011; Nudler, 2012).

La formation de R-Loops est défavorisée par la présence de ribosomes sur le transcrit naissant, probablement car ceux-ci piègent le transcrit ARN (empêchant sa ré-hybridation avec le brin ADN matrice) et/ou empêchent la rétro-translocation de l’ARNP (Nudler, 2012; Proshkin *et al.*, 2010). A l’inverse, les R-Loops sont préférentiellement formées quand l’ARN est nu (Masse & Drolet, 1999b), ce qui en fait des substrats potentiels pour le facteur Rho. Récemment, il a été montré que la formation de R-Loops est favorisée dans un mutant *nusG* d’*E. coli* (Leela *et al.*, 2013). L’analyse de la distribution de ces R-Loops montre qu’elles sont préférentiellement présentes à des loci du génome où la transcription et la traduction sont susceptibles de ne pas être couplées (transcription de longs opérons plus facilement sujets à des effets de polarité transcriptionnelle, transcription antisense, transcription de régions inter-géniques). De plus, les auteurs indiquent que l’inactivation d’EcRho (par délétion du gène *rho*), habituellement létale, peut être compensée par l’expression hétérologue d’UvsW, une hélicase du phage T4 impliquée dans la dislocation d’hétéroduplexes viraux (Leela *et al.*,

2013). Ces observations amènent les auteurs à conclure que la résolution des R-Loops serait la fonction majeure de Rho (Leela *et al.*, 2013).

Un mécanisme différent par lequel EcRho contribue plus modestement à la stabilité du génome a été découvert récemment. Dans ce cas, EcRho inhibe l'expression d'un gène spécifique, codant la protéine RDF (*Recombination Directionality Factor*), nécessaire à l'excision de prophages dont la présence est bénéfique pour *E. coli* (Menouni *et al.*, 2013).

De l'ensemble de ces observations, il apparaît que le facteur Rho est impliqué dans un grand nombre de fonctions biologiques importantes. Cette diversité explique très probablement pourquoi Rho est une enzyme essentielle pour *E. coli*, espèce chez qui la plupart de ces fonctions ont été découvertes. Rho pourrait représenter une cible thérapeutique particulièrement intéressante si ces fonctions étaient conservées chez une majorité d'espèces bactériennes, d'autant que Rho semble être absent chez les eucaryotes.

5. RHO ET LA DIVERSITÉ BACTÉRIENNE

A. Spectre de la bicyclomycine et importance de Rho

La bicyclomycine (BCM) est un antibiotique produit par différentes souches de *Streptomyces* isolées du sol japonais (Miyamura *et al.*, 1972; Miyoshi *et al.*, 1972; Nishida *et al.*, 1972) et reste à ce jour le seul inhibiteur naturel connu qui est spécifiquement dirigé contre le facteur Rho (Zwiefka *et al.*, 1993). Ce composé hétérocyclique (**Figure 43.A**), dont la biosynthèse est dérivée de celles de la leucine et de l'isoleucine (Iseki *et al.*, 1980; Miyoshi *et al.*, 1980), est un inhibiteur non compétitif de l'activité ATPase de Rho (Park *et al.*, 1995). Certaines données biochimiques suggèrent que la fixation de la BCM a également une influence sur la capacité du SBS à interagir avec l'ARN (Magyar *et al.*, 1996). Plusieurs structures cristallines du facteur EcRho en interaction avec la BCM (ou des dérivés synthétiques) ont été publiées (Skordalakes *et al.*, 2005). Dans ces structures, l'hexamère EcRho est en conformation ouverte avec chaque poche ATPase occupée par une molécule d'ATP γ S (**Figure 43.B**). La fixation de la BCM dans un site adjacent de la poche ATPase se situe près du glutamate catalytique (Glu211; **Figure 43.B**) et obstrue l'espace occupé normalement par la molécule d'H₂O attaquant le groupement γ -phosphoryle de l'ATP (Balasubramanian & Stitt, 2010; Skordalakes *et al.*, 2005; Thomsen & Berger, 2009). Les auteurs observent que la BCM forme des interactions avec le phosphate de l'ATP γ S qui

pourraient également interférer avec l'hydrolyse et le relargage du nucléotide (Skordalakes *et al.*, 2005). Enfin, la BCM forme des interactions avec le motif R-Loop d'EcRho (Skordalakes *et al.*, 2005), ce qui pourrait expliquer l'effet qu'a la drogue sur la fixation du SBS à l'ARN (**Figure 43.B**) (Magyar *et al.*, 1996).

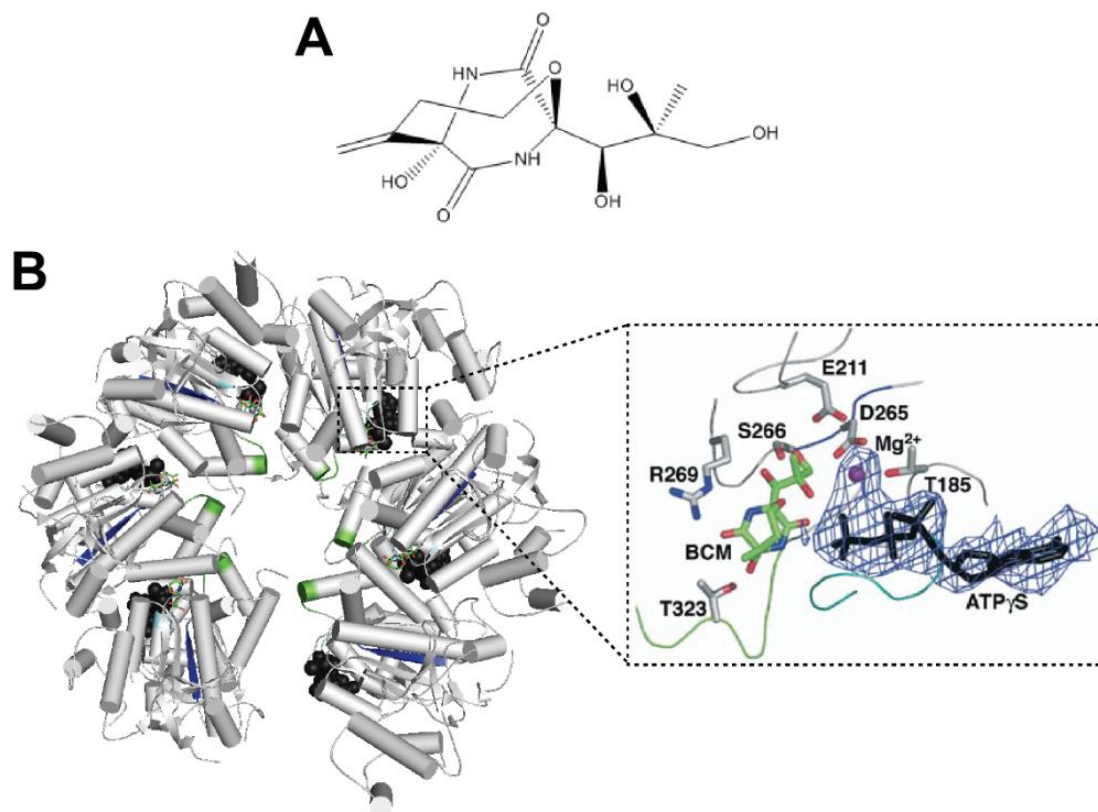


Figure 43 : Mécanisme d'inhibition de la bicyclomycine. (A) Représentation détaillée de la BCM. (B) Structure cristalline d'EcRho en association avec la BCM et l'ATP γ S (PDB : 1XPO). Les molécules d'ATP γ S sont représentées en noir (adapté de Skordalakes *et al.*, 2005).

La sensibilité à la BCM a souvent été utilisée comme critère du caractère essentiel de Rho chez les bactéries (Godoy *et al.*, 2008; Miyoshi *et al.*, 1972; Nishida *et al.*, 1972). Ainsi, la BCM inhibe la croissance de nombreuses espèces Gram⁻ alors que la plupart des espèces Gram⁺ testées, à l'exception de *Micrococcus luteus*, sont résistantes (**Tableau 3**). Cependant, les espèces testées sont loin de refléter la diversité bactérienne. Les représentants Gram⁻ appartiennent principalement au super-phylum des Protéobactéries (**Tableau 3**) et l'on peut noter l'absence d'espèces appartenant à d'autres phyla Gram⁻ (Cyanobactéries, Chlamydiae, Planctomycetes, ...). Un constat similaire peut être fait pour les Gram⁺ où la pluparts des représentants n'appartiennent qu'aux phyla des Firmicutes et des Actinobacteries (**Tableau 3**). L'idée que Rho est une enzyme essentielle principalement chez les Gram⁻ doit donc être considérée avec précaution.

Tableau 3 : Concentrations minimales inhibitrices (CMI) de BCM mesurées pour certaines espèces bactériennes ^a

Phylum	Espèce	CMI (µg/mL)
α-Protéobactéries	<i>Caulobacter crescentus</i>	200
β-Protéobactéries	<i>Neisseria gonorrhoeae</i>	25
	<i>Neisseria meningitidis</i>	> 800
γ-Protéobactéries	<i>Escherichia coli</i>	25
	<i>Klebsiella pneumoniae</i>	100
	<i>Shigella flexneri</i>	de 12,5 à 25 ^b
	<i>Salmonella enterica</i>	25
	<i>Aeromonas sp.</i>	de 1 à 100 ^b
ε-Protéobactéries	<i>Campylobacter ureolyticus</i>	de 0,5 à >1000 ^b
Spirochètes	<i>Brachyspira hyodysenteriae</i>	>100
Bactéroidetes	<i>Bacteroides thetaiotaomicron</i>	de 32 à 512 ^b
	<i>Bacteroides fragilis</i>	de 2 à 1000 ^b
Firmicutes	<i>Staphylococcus aureus</i>	> 800
	<i>Bacillus subtilis</i>	> 800
	<i>Streptococcus pneumoniae</i>	> 800
	<i>Enterococcus faecalis</i>	> 800
	<i>Clostridium perfringens</i>	1000
Actinobactéries	<i>Micrococcus luteus</i>	250
	<i>Corynebacterium diphtheriae</i>	≥ 800
	<i>Mycobacterium tuberculosis</i>	> 800
	<i>Streptomyces lividans</i>	> 600
	<i>Streptomyces coelicolor</i>	> 600

^a d'après (Godoy *et al.*, 2008; Italiani & Marques, 2005; Kitai *et al.*, 1979; Nishida *et al.*, 1972; Watt & Brown, 1983).

^b selon les souches.

B. La distribution du facteur Rho au sein de la diversité bactérienne

Seules deux études relativement anciennes se sont intéressées à la distribution de gènes codant pour une protéine Rho putative au sein de la biodiversité (Opperman & Richardson, 1994; Washburn *et al.*, 2001). Limitées par le faible nombre de séquences disponibles à l'époque, ces études suggèrent néanmoins que de nombreuses espèces bactériennes possèdent un gène codant une protéine Rho, à l'exception de quelques Firmicutes et de la Cyanobactérie *Synechocystis sp.* PCC6803 (Washburn *et al.*, 2001). Ces études suggèrent également que Rho est une protéine dont la séquence est relativement conservée, notamment au niveau du CTD et des motifs clés Walker A, Walker B, RNP1 et RNP2 (**Figure 29.A & Figure 44**). Il a cependant été noté que le facteur Rho de *M. luteus*

(MlutRho) possède un domaine NTD sensiblement plus grand qu'EcRho (Nowatzke & Richardson, 1996; Opperman & Richardson, 1994). Cette différence est due à la présence d'une insertion de 263 acide-aminés située entre les motifs NHB et CSD-Like (**Figure 44**) (Nowatzke & Richardson, 1996). Depuis, des insertions NTD similaires ont été retrouvées chez d'autres facteurs Rho issus d'espèces appartenant au phylum des Actinobactéries (**Figure 44**) (Ingham *et al.*, 1996; Kalarickal *et al.*, 2010; Nowatzke & Richardson, 1996). Ces insertions n'ont aucune relation phylogénique, aussi bien au niveau de la conservation de séquences qu'au niveau de leurs tailles (voir Chapitre IV). Leur unique point commun est leur richesse en résidus polaires (notamment en Arg, Asp, Lys et Gly) suggérant qu'elles sont non-structurées. La présence de ces insertions dans les domaines NTD permettrait l'association du facteur Rho avec des ARN possédant un taux de structures secondaires élevées, en relation avec le haut taux de G/C retrouvé dans le génome des Actinobactéries (Nowatzke *et al.*, 1997a).

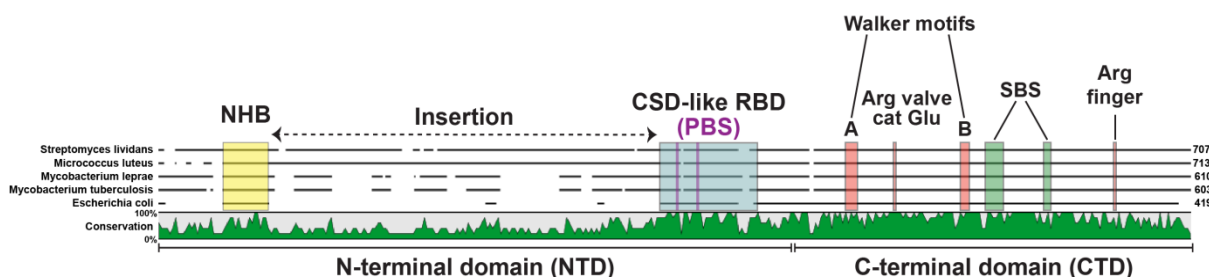


Figure 44 : Alignement des séquences d'EcRho et de facteurs Rho d'Actinobactéries. Les motifs conservés sont représentés avec le même code couleur que dans la Figure 29. Le graphe (en vert) montre les taux de conservation des acides aminés.

Bien que l'on ne retrouve pas de gène codant pour *rho* dans le génome des eucaryotes (voir Article I, page 114), la terminaison de la transcription médiée par une hélicase est un mécanisme que l'on retrouve chez *S. cerevisiae* et *Arabidopsis thaliana* via les protéines Sen1 et RHON1, respectivement (Chi *et al.*, 2014; Porrua & Libri, 2013). Chez *S. cerevisiae*, Sen1 fait partie du complexe NNS prédit pour induire la terminaison de la transcription d'ARNnc, limiter la transcription antisense et inhiber la formation de R-Loops (Chi *et al.*, 2014; Mischo *et al.*, 2011; Porrua & Libri, 2013; Porrua & Libri, 2015). Chez *A. thaliana*, RHON1 participe à la terminaison de gènes plastidiques (Chi *et al.*, 2014). Cependant, aucun mécanisme moléculaire détaillé de la terminaison de la transcription n'a encore été défini pour ces deux enzymes. Chez l'homme, la senataxin (SETX), un homologue de Sen1, a été impliquée dans la résolution des R-Loops (Bennett & La Spada, 2015).

C. La conservation des mécanismes de Rho

Chez certaines espèces bactériennes, le facteur Rho semble utiliser des mécanismes différents de ceux mis en évidence pour *E. coli*, notamment pour la reconnaissance des sites *Rut*. Par exemple, le facteur Rho de *Rhodobacter sphaeroides* (RspRho), une α -Protéobactérie dont le génome est riche en G/C, ne semble pas capable de reconnaître les sites *Rut* naturels d'*E. coli* et ne présente pas d'activité ATPase en présence de poly[rU] (Ingham, 1999). En caractérisant l'affinité d'une protéine chimère composée du domaine NTD de RspRho et du domaine CTD d'EcRho, les auteurs ont pu mettre en évidence que ce défaut de reconnaissance était dû à la divergence de séquences de leurs domaines NTD respectifs (Ingham, 1999).

Tous les facteurs Rho n'hydrolysent pas les NTP de la même manière, certains présentant une spécificité de substrat plus marquée. Par exemple, les facteurs Rho d'Actinobactéries semblent hydrolyser préférentiellement les purines triphosphates (Ingham *et al.*, 1996; Mitra *et al.*, 2014; Nowatzke & Richardson, 1996). De plus, il a été rapporté que le facteur Rho de *M. luteus* est capable d'induire une terminaison précoce de la transcription à des sites non reconnus par EcRho (Nowatzke *et al.*, 1997a). Cette propriété est de nouveau liée à la divergence des domaines NTD et, dans ce cas précis, à la présence de l'insertion NTD de 263 acides aminés dans MlutRho (voir ci-dessus). En effet, cette insertion semble conférer à MlutRho la capacité de reconnaître et de se fixer à des sites *Rut* plus structurés que ceux que peut utiliser EcRho (Nowatzke *et al.*, 1997a).

Une terminaison précoce de la transcription a également été démontrée pour le facteur Rho de *M. tuberculosis* (MtbRho), une Actinobactérie dont le génome est lui aussi riche en G/C (Kalarickal *et al.*, 2010). Les auteurs de l'étude ont fait d'autres observations intrigantes : MtbRho aurait une activité ATPase beaucoup plus faible qu'EcRho et serait totalement dépourvue d'activité hélicase (Kalarickal *et al.*, 2010). Les auteurs proposent que MtbRho est un facteur possédant un faible moteur moléculaire ce qui ne l'empêche pourtant pas d'avoir une activité robuste de terminaison de la transcription. De plus, bien que la BCM inhibe l'activité ATPase de MtbRho, l'enzyme reste capable de provoquer la dissociation de CET artificiellement stoppés le long de l'ADN en présence de BCM (Kalarickal *et al.*, 2010). Mais le plus surprenant réside dans le fait que MtbRho est capable de dissocier ces CET artificiellement stoppés, même lorsque le cofacteur ATP est remplacé par un analogue non hydrolysable (ATP γ S ou AMP-PNP). Ces observations conduisent les auteurs à proposer que

le facteur MtbRho est capable de provoquer la terminaison de la transcription par un mécanisme qui est indépendant de son moteur moléculaire (Kalarickal *et al.*, 2010). Ce mécanisme indépendant de l'hydrolyse de NTP, s'il était répandu au sein de la diversité bactérienne, pourrait contribuer à expliquer le spectre restreint de la BCM. Par ailleurs, l'éventualité que certains facteurs Rho phylo-divergents puissent provoquer la dissociation du CET par un mécanisme différent de celui utilisé par EcRho, ouvre un nouveau champ d'investigation dans l'étude de la transcription bactérienne.

Chapitre IV :

Résultats personnels

1. INTRODUCTION

Ma thèse porte sur l'étude du facteur Rho, une hélicase hexamérique qui est capable de reconnaître un site *Rut* et de se déplacer de manière ATP-dépendante le long du transcrit ARN pour rattraper l'ARNP et provoquer la terminaison de la transcription chez les bactéries (**Figure 1**). Aujourd'hui, même si certains aspects restent encore mal définis ou sujets à débats, les mécanismes moléculaires gouvernant la terminaison Rho-dépendante chez *E. coli* sont relativement bien compris (Boudvillain *et al.*, 2010; Peters *et al.*, 2011; Washburn & Gottesman, 2015). De plus, l'utilisation croissante de techniques d'analyse à l'échelle génomique/transcriptomique a mis en évidence ou confirmé l'importance d'EcRho dans diverses fonctions biologiques clés (Cardinale *et al.*, 2008; Leela *et al.*, 2013; Mooney *et al.*, 2009a; Peters *et al.*, 2012). Toutes ces données nous permettent maintenant de mieux comprendre pourquoi Rho est une enzyme essentielle chez *E. coli* et les espèces apparentées. Cependant, l'importance, le rôle et les mécanismes du facteur Rho dans la diversité bactérienne sont, par comparaison, mal documentés. La BCM, un antibiotique inhibant l'activité ATPase de Rho, semble être efficace principalement contre les Protéobactéries mais sans effet sur la plupart des espèces Gram⁺ testées. Durant ma thèse, je me suis efforcé de mieux comprendre ce spectre d'activité et de mieux cerner le périmètre de Rho dans la biodiversité en testant deux hypothèses principales:

1) Est-il possible que de nombreuses espèces bactériennes, en particulier des Gram⁺, ne possèdent pas de gène codant pour un facteur Rho ?

Les études de la distribution et de la conservation du facteur Rho restaient très limitées malgré l'avènement de l'ère génomique et l'expansion faramineuse des banques de données biologiques. Nous avons exploité cette richesse récemment accumulée pour réévaluer l'influence de Rho dans la biodiversité. Nous avons déterminé que 1250 des 1349 espèces bactériennes (>92%) dont le génome avait été séquencé au début de l'étude (début 2012) possèdent un gène *rho*. Malgré cette large représentation, *rho* semble être totalement absent de quelques phyla (Cyanobactérie et Ténéricutes) et de quelques sous-embranchements des Firmicutes (la classe des *Negativicutes* et l'ordre des *Lactobacilliales*). La comparaison des génomes contenant *rho* avec ceux qui en sont dépourvus suggère que la présence de Rho est liée à la complexité de l'organisation et/ou de l'expression génomique, en relation avec le style de vie de la bactérie. Nous avons aussi observé que certaines espèces possèdent deux copies du gène *rho*, ce qui semble parfois dû à un évènement de duplication du gène ancestral

mais reste d'origine indéterminée (transfert horizontal ?) dans de nombreux cas. En alignant et comparant l'ensemble de nos séquences Rho putatives (1259 séquences après nettoyage de l'échantillon), nous avons observé qu'environ un tiers d'entre elles, majoritairement issues des phyla des Actinobactéries et des Bactéroidetes, présentaient de larges insertions dans leurs NTD. De plus, une proportion non-négligeable (~ 5%) des séquences Rho contient des mutations connues pour diminuer l'affinité pour la BCM, ce qui pourrait contribuer à expliquer le spectre d'action restreint de la drogue. Ces résultats ont fait l'objet d'une publication dans le journal *Microbiology* en 2013 (**Article I**, page 109).

2) Chez certaines espèces, le facteur Rho utilise-t-il un mécanisme atypique, indépendant de l'hydrolyse d'ATP ?

Ceci a été proposé pour le facteur MtbRho de *M. tuberculosis* qui semble être un moteur moléculaire peu actif et incapable de dissocier un hybride ARN:ADN de 20 pb artificiellement formé par appariement d'un oligonucléotide avec la région *RutB* d'un transcrit λ tR1 naissant (Kalarickal *et al.*, 2010). Cette absence d'activité hélicase apparente n'empêche pourtant pas MtbRho d'induire une terminaison de la transcription robuste et précoce, y compris en présence de quantités de BCM qui sont inhibitrices pour EcRho (Kalarickal *et al.*, 2010). De plus, MtbRho est capable de dissocier un CET artificiellement stoppé le long de l'ADN, même lorsque le seul nucléotide cofacteur présent est l'ATP γ S ou l'AMP-PNP (analogues considérés comme non-hydrolysables) (Kalarickal *et al.*, 2010). Ces caractéristiques rappellent le cas de certaines ARN hélicases DEAD-box qui sont capables de détruire de courts duplexes ARN de manière non directionnelle en présence d'analogues non-hydrolysables de l'ATP (voir page 58) (Liu *et al.*, 2008). Il est proposé dans ce cas que l'association coopérative du nucléotide cofacteur et de l'ARN à l'hélicase DEAD-box est suffisamment forte pour déformer et déstabiliser la courte région en duplexe (Chen *et al.*, 2008; Hilbert *et al.*, 2009; Liu *et al.*, 2008; Mallam *et al.*, 2012). Nous avons testé si MtbRho pouvait utiliser un mécanisme similaire pour provoquer la dissociation du CET. Pour cela, nous avons étudié en détail les propriétés enzymatiques de la forme sauvage de MtbRho et de diverses formes variantes (mutants catalytiques, mutants de délétions, chimère MtbRho/EcRho). Nos résultats confirment que l'activité ATPase de MtbRho est faible, ce qui ne l'empêche pas d'induire une terminaison de la transcription forte et précoce. Nous montrons, en revanche, que MtbRho possède une activité hélicase directionnelle (5'→3'), comme EcRho, même si cette activité ne peut s'exercer que sur de courts duplexes. Cette faible processivité apparente est en partie conférée par la présence d'une thréonine à une

position clé du SBS (au lieu d'une lysine chez la plupart des autres espèces) et par la présence d'une insertion de 143 acides aminés dans la partie NTD de l'enzyme. De plus, nous démontrons que les activités hélicase et de terminaison de la transcription de MtbRho sont toutes les deux strictement dépendantes de la capacité de l'enzyme à hydrolyser l'ATP et qu'elles sont inhibées à de fortes concentrations de BCM. Ces résultats réfutent l'hypothèse d'un nouveau mécanisme de type DEAD-box et suggèrent que le mécanisme mis en avant pour EcRho est globalement conservé dans le règne bactérien malgré quelques variations pouvant être liées aux caractéristiques propres (composition du génome, vitalité du métabolisme, etc.) de certaines espèces. Ces résultats ont fait l'objet d'une publication dans le journal *Nucleic Acids Research* en 2015 (**Article II**, page 131).

En plus de l'exploration de ces deux aspects généraux de la fonction de Rho, j'ai eu l'opportunité de participer à une étude dédiée à la mise en évidence d'un nouveau mécanisme de régulation Rho-dépendante. Dans le cadre d'une collaboration étroite avec l'équipe de Nara Figueroa-Bossi (CNRS, Gif/Yvette), nous avons ré-analysé la liste de loci Rho-dépendants révélés par analyses génomiques et transcriptomiques chez *E. coli* (Peters *et al.*, 2009; Peters *et al.*, 2012), à la recherche d'exemples potentiels de mécanismes d'atténuation Rho-dépendante dans la région 5'-leader de l'ARNm. Cette analyse a identifié un candidat particulièrement intéressant, l'opéron *pgaABCD*, impliqué dans la formation de biofilms chez *E. coli* (Wang *et al.*, 2004). La régulation de *pgaABCD* est relativement complexe, avec des composantes transcriptionnelles et post-transcriptionnelles impliquant notamment un sRNA (McaS) et la protéine CsrA. Cette dernière est un facteur de régulation capable d'interagir avec une fraction importante des ARNm d'*E. coli* (> 20%) (Holmqvist & Vogel, 2013; Romeo *et al.*, 2013). Des sites d'interaction pour McaS (Thomason *et al.*, 2012) et CsrA (Wang *et al.*, 2005) avaient été identifiés dans la longue (234 nt) région 5'-leader de l'ARNm *pgaABCD*. Nous avons observé que ces sites d'interactions se trouvent dans, ou à proximité, d'une séquence possédant les attributs d'un site *Rut* potentiel (bulle C>G, groupes de dinucléotides 5'-YC correctement espacés). Nous avons alors démontré qu'en accord avec cette prédiction la région 5' UTR de *pgaABCD* contenait bien un terminateur Rho-dépendant fonctionnel *in vitro*. Néanmoins, l'efficacité de terminaison est très dépendante de la force ionique *in vitro* et requière la présence de NusG (qui stimule les terminateurs faibles; voir page 84) ou de CsrA à des concentrations salines modérées à élevées (≥ 150 mM KCl). La présence d'un terminateur Rho-dépendant sous le contrôle de CsrA (et NusG) a été validée *in vivo* à l'aide de systèmes rapporteurs développés par nos collaborateurs chez *Salmonella*. Ces

systèmes rapporteurs ont également permis de montrer, en accord avec un article publié pendant notre étude (Jørgensen *et al.*, 2013), que l'action de McaS s'exerçait indirectement, en séquestrant CsrA (et non en se liant au 5'-leader de *pgaABCD*). Enfin, en combinant analyses *in vitro* et *in vivo*, nous avons démontré qu'en se fixant à la région 5'-leader de *pgaABCD*, CsrA détruisait une structure secondaire qui séquestre une partie du site *Rut* et rend ce dernier inactif. Ces résultats révèlent un nouveau mécanisme de régulation génique de type "riboswitch", où l'accessibilité d'un terminateur Rho-dépendant est modulée par la protéine CsrA. Ces résultats ont fait l'objet d'une publication dans le journal *Genes & Development* en 2014 (**Article III**, page 151).

ARTICLE I:

**PHYLETIC DISTRIBUTION AND CONSERVATION OF THE
BACTERIAL TRANSCRIPTION TERMINATION FACTOR RHO**

Microbiology (2013)

Phyletic distribution and conservation of the bacterial transcription termination factor Rho

François D'Heygère,^{1,2} Makhlof Rabhi¹ and Marc Boudvillain^{1,3}

Correspondence

Marc Boudvillain

marc.boudvillain@cnrs-orleans.fr

¹Centre de Biophysique Moléculaire, CNRS UPR4301, rue Charles Sadron, 45071 Orléans cedex 2, France

²Ecole doctorale Santé, Sciences Biologiques et Chimie du Vivant (ED 549), Université d'Orléans, France

³ITP Sciences Biologiques et Chimie du Vivant, Université d'Orléans, France

Transcription termination factor Rho is a ring-shaped, ATP-dependent molecular motor that targets hundreds of transcription units in *Escherichia coli*. Interest in Rho was renewed recently on the realization that this essential factor is involved in multiple interactions and cellular processes that protect the *E. coli* genome and regulate its expression on a global scale. Yet it is currently unknown if (and how) Rho-dependent mechanisms are conserved throughout the bacterial kingdom. Here, we mined public databases to assess the distribution, expression and structural conservation of Rho across bacterial phyla. We found that *rho* is present in more than 90 % of sequenced bacterial genomes, although Cyanobacteria, Mollicutes and a fraction of Firmicutes are totally devoid of *rho*. Genomes lacking *rho* tend to be small and AT-rich and often belong to species with parasitic/symbiotic lifestyles (such as Mollicutes). By contrast, large GC-rich genomes, such as those of Actinobacteria, often contain *rho* duplicates and/or encode Rho proteins that bear insertion domains of unknown function(s). Notwithstanding, most Rho sequences across taxa contain canonical RNA-binding and ATP hydrolysis signature motifs, a feature suggestive of largely conserved mechanism(s) of action. Mutations that impair binding of bicyclomycin are present in ~5 % of *rho* sequences, implying that species from diverse ecosystems have developed resistance against this natural antibiotic. Altogether, these findings assert that Rho function is widespread among bacteria and suggest that it plays a particularly relevant role in the expression of complex genomes and/or bacterial adaptation to changing environments.

Received 1 March 2013

Accepted 15 May 2013

INTRODUCTION

In bacteria, the Rho factor mediates one of the two major pathways that lead to the termination of transcription (Ciampi, 2006; Peters *et al.*, 2011). Since eukaryotes do not contain Rho homologues, the protein makes an attractive target for the development of novel antibiotics. Early studies with the natural Rho inhibitor bicyclomycin (BCM) delineated the spectrum of organisms susceptible to Rho inhibition (Nishida *et al.*, 1972). BCM is active against various Gram-negative bacteria, but inactive against all tested Gram-positive bacteria with the exception of *Micrococcus (Mc.) luteus* (Nowatzke *et al.*, 1997b; Williams & Durham, 1988). This has been interpreted to

suggest that Rho function is important in Gram-negative bacteria, but dispensable in most Gram-positive species (Rabhi *et al.*, 2010b; Washburn *et al.*, 2001). The rationale for this simple rule is, however, not firmly established. Notably, the phyletic distribution and structural conservation of Rho within living organisms were inferred more than a decade ago from very limited sets of DNA sequences (Opperman & Richardson, 1994; Washburn *et al.*, 2001). Hence, the fractions of bacteria devoid of *rho* genes or containing Rho factors whose function may be unaffected by BCM are not known.

Typically, the Rho factor is a ring-shaped homo-hexameric enzyme with RNA-dependent ATP hydrolase activity (reviewed by Banerjee *et al.* 2006; Boudvillain *et al.*, 2010; Ciampi, 2006; Peters *et al.*, 2011). This ATPase activity fuels the translocation of Rho along nascent transcripts and the ensuing displacement of roadblocks formed by RNA:DNA duplexes (e.g. transcriptional R-loops) or by RNA polymerases halted at Rho-dependent termination sites (Fig. 1a). Rho function is tightly controlled *in vivo* and

Abbreviations: BCM, bicyclomycin; CDS, coding sequence; CSD, cold shock domain; CTD, C-terminal domain; HGT, horizontal gene transfer; NHB, N-terminal helix bundle; NTD, N-terminal domain; PBS, primary binding site; RBD, RNA-binding domain; SBS, secondary binding site.

Four supplementary figures and four supplementary tables are available with the online version of this paper.

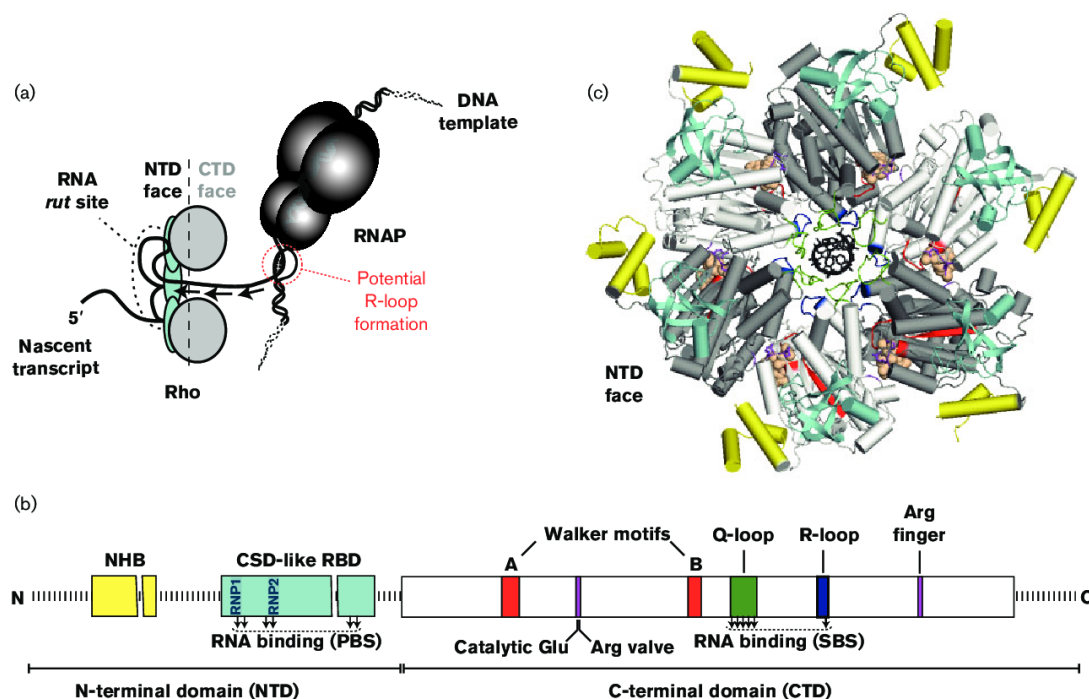


Fig. 1. Transcription termination factor Rho. (a) Schematic of Rho-dependent termination (individual components are not set to scale). (b) Diagram depicting the arrangement of key motifs within the primary sequence of the Rho peptide. Regions of variable length are represented by dotted lines. Arrows represent amino acids contacting RNA in structures of ecRho (PDB no. 3ICE and no. 1PVO). (c) Crystal structure of the Rho hexamer from *E. coli* (PDB no. 3ICE). Important domains are coloured as in (b). Other regions are alternatively depicted in white or grey in adjacent monomers in order to make monomer interfaces more discernible. The ADP-BeF₃ ligands present in the ATPase pockets are depicted in light brown space-filled balls. A short r(U)₆ fragment bound to the central channel of the hexamer is shown in black. Note that no RNA binds Rho's PBS in this structure.

usually requires uncoupling of translation from transcription (or lack of translation), which exposes nascent RNA *rut* (Rho utilization) sites to binding by Rho. Activation of Rho function is nonetheless frequent in *Escherichia coli* where Rho mediates many transcription termination events (current estimates, 20–50 %) and regulates hundreds of transcription units encoding proteins as well as non-coding RNAs (transfer RNAs, small RNAs and antisense RNAs) (Cardinale *et al.*, 2008; Peters *et al.*, 2009, 2012). Accordingly, diverse biological roles were proposed for Rho in *E. coli* and related bacteria (reviewed by Boudvillain *et al.*, 2013). They include control of faithful gene expression [e.g. by transcriptional polarity (Ciampi, 2006; Peters *et al.*, 2011)], regulation of specific genes/operons in response to environmental cues [notably through ribo-switch- and small RNA-dependent mechanisms (Bossi *et al.*, 2012; Hollands *et al.*, 2012)], prevention of spurious transcription (Cardinale *et al.*, 2008; Peters *et al.*, 2009, 2012) and of R-loop formation (Krishna Leela *et al.*, 2013), and roles in maintenance of genome integrity (Dutta *et al.*, 2011; Washburn & Gottesman, 2011). These functions rely

more-or-less directly on Rho cooperating with various endogenous factors. Notably, Rho binds to transcription factors NusA and NusG (Li *et al.*, 1993; Schmidt & Chamberlin, 1984), antitermination factor YaeO (Pichoff *et al.*, 1998) and riboregulation factor Hfq (Rabhi *et al.*, 2011b), and co-purifies with a host of other factors (Butland *et al.*, 2005; Hu *et al.*, 2009) that vary with growth conditions to include, for instance, RNA degradosome components PNPase and RNase E (Tuckerman *et al.*, 2011). At present, the extent to which these various roles and interactions are conserved throughout the bacterial kingdom remains largely uncertain. This is due to past research that focused primarily on Rho from *E. coli* and on very few other bacteria. For instance, the possibility that some Rho homologues operate by distinctive mechanisms is supported by only one recent *in vitro* study devoted to the Rho factor from *Mycobacterium tuberculosis* (Kalarickal *et al.*, 2010). The authors show that the transcription termination activity of *Mb. tuberculosis* Rho can be totally uncoupled from ATP hydrolysis, an unexpected feature that may be linked tentatively to the presence of a large

N-terminal domain (NTD) insertion within the Rho polypeptide (Kalarickal *et al.*, 2010). Since *rho* inactivation (Griffin *et al.*, 2011), but not BCM treatment (Nishida *et al.*, 1972), inhibits mycobacterial growth, the data also argue that sensitivity to the ATPase inhibitor BCM (Skordalakes *et al.*, 2005) is not always a reliable gauge of the functional significance of Rho.

To better appraise the importance of Rho across the bacterial kingdom, we explored public databases for clues supporting the presence and functional role of *rho* in the thousands of bacterial genomes that have now been sequenced. We also assessed the conservation of important residues/motifs revealed by previous studies of Rho, including (1) the cold shock domain (CSD)-like RNA-binding domain (RBD) that is located within Rho's NTD and ensures recruitment of transcript *rut* regions on the crown-like primary binding site (PBS) of the Rho hexamer (Fig. 1b, c, in cyan), (2) the positively charged N-terminal helix bundle (NHB) domain located on the periphery of the hexamer (Fig. 1b, c, in yellow) that may facilitate RNA recruitment to the PBS (Canals *et al.*, 2010), (3) ATP hydrolysis signature motifs [Walker A (P-loop) and B motifs] located within Rho's C-terminal domain (CTD) (Fig. 1b, c, in red), (4) CTD residues involved in coordinating ATP hydrolysis or structural transitions within the hexamer [e.g. 'catalytic' Glu, Arg 'valve' and Arg 'finger' residues (Fig. 1b, c, in purple)] and (5) CTD residues forming the secondary binding site (SBS) (Q- and R-loops, in green and blue in Fig. 1b, c, respectively) and required for RNA translocation within the ring channel (Rabhi *et al.*, 2010a). The arrangement of these components in a crystal structure of *E. coli* Rho (*ecRho*), which shows the hexamer in a closed, asymmetrical configuration (Fig. 1c), suggests a mechanism for RNA translocation, whereby each nucleotide is SBS-bound by a single Rho subunit and driven through the central channel in response to sequential chemical changes in the ATP-binding pockets (Thomsen & Berger, 2009). However, this monotonous, ATPase-driven 'escort' mechanism does not account for the ~7 nt periodic stepping behaviour of *ecRho* (Boudvillain *et al.*, 2010; Schwartz *et al.*, 2009) nor explain how Rho from *Mb. tuberculosis* triggers efficient transcription termination in spite of very poor ATPase activity (Kalarickal *et al.*, 2010). Thus, we have also looked for features that could reflect alternative Rho mechanisms, notably the presence of atypical domain insertions and/or aa substitutions. We show that the Rho factor is generally well conserved and expressed across the bacterial kingdom, but highlight intriguing features in a number of cases that will require further exploration and, eventually, experimental assessment. We observe that *rho* is absent from a limited number of species (~8 % of our compiled dataset), which include all Cyanobacteria and Mollicutes as well as a significant fraction of Firmicutes. Species from various phyla were found to have duplicated and/or atypical *rho* genes, including ones bearing deletions/insertions of unknown function(s) or mutations that are likely to alter

enzymic activity or to confer resistance to BCM. Interestingly, genomes containing *rho* sequence(s) tend to be larger and GC-richer than *rho*-less genomes, suggesting a link between Rho function and the complexity of genome structure (or expression) and/or lifestyle of the bacterial species. Overall, our work significantly extends early phylogenetic analyses (Opperman & Richardson, 1994; Washburn *et al.*, 2001) that were performed with limited sets of bacterial sequences and before detailed structural information on Rho became available, thereby providing a comprehensive framework for future analyses of this central regulatory factor.

METHODS

Database searches. We searched the Xbase, Uniprot, GenBank and NCBI protein databases for *rho* coding sequences (CDSs) or ORFs and for Rho proteins using a combination of keyword queries, as well as BLAST searches with selected target sequences. Sequences that displayed homology with Rho NTD (PSSM Id: 208474) only, such as hypothetical protein MAE_06000 (414 aa) from *Microcystis aeruginosa*, were not considered significant hits. We thoroughly checked all available genomes of species lacking *rho* to ensure absence of the gene itself or lack of any significant remnant. Control searches were conducted with PSI-BLAST (threshold *E*-value=0.05) and distinct Rho baits, including ones from bacteria having phyletic proximity with the *rho*-less species (proximity based on SILVA's tree of life; <http://www.arb-silva.de>). Rho baits were b3783 (*E. coli*), Rv1297 (*Mb. tuberculosis*), BSU37080 (*Bacillus subtilis*), HMPREF9087_1232 (*Enterococcus casseliflavus*), sce7932 (*Sorangium cellulosum*) and Hprae_1978 (*Halanaerobium praevalens*). Consistent with long-known homology (Opperman & Richardson, 1994), usually the best PSI-BLAST hits for *rho*-less species were ATP synthase subunits with 40–60 % coverage (corresponding to Rho's CTD) and 20–30 % identity scores.

When several completed genomes existed for a given species, we arbitrarily selected the strain with the largest genome or the most representative strain (e.g. *E. coli* K-12 MG1655) to build our datasets. Whenever possible, we also favoured fully assembled genomes over draft or incomplete ones. To perform all subsequent analyses on uniform datasets, searches for *rho*⁺ and *rho*-less species were stopped at the end of January 2012. Locus tags for the annotated *rho* genes were used to retrieve Rho sequences from the Uniprot database and to build FASTA files. Theoretical isoelectric points and hydrophobicity GRAVY (grand average of hydropathicity) indexes were calculated with PROTPARAM on the ExPASy portal (<http://www.expasy.org>).

Sequence alignments and phylogeny trees. Sequence alignments and phylogeny trees were constructed with SEAVIEW 4.3.3 (Gouy *et al.*, 2010). Sequences were aligned with the MUSCLE algorithm implemented in SEAVIEW using default parameters. Unrooted phylogeny trees were built with the BioNJ distance method (bootstrap=100; gap sites ignored). Iterative cycles of sequence alignment and phylogenetic ordering were performed until convergence of both was obtained. Detection and alignment of NHB and NTD insertion domain residues over the full dataset were more satisfactory with this approach than by direct alignment with alternative algorithms such as CLUSTAL-Ω (not shown). Phylogeny was then evaluated with maximum-likelihood (PHYML) and/or parsimony (PHYLP) algorithms, as implemented in SEAVIEW. Bootstrap support was estimated based on 100 replicates. Conversion of alignments into image files was performed with JALVIEW (Waterhouse *et al.*, 2009) using the FASTA files generated with SEAVIEW. Histograms for sequence conservation (based on physicochemical properties), quality (unlikelihood to observe a

non-conservative substitution in the alignment column) and consensus (percentage of consensus nucleotide in the column) were generated with the default parameters implemented in JALVIEW. Conservation and quality histograms are shaded by default from dark brown (low values) to yellow (high values). Shannon entropy analysis was performed on the Protein Variability Server (Garcia-Boronat *et al.*, 2008).

RESULTS AND DISCUSSION

Rho homologues are found in many bacterial phyla

Database mining provided evidence for a *rho* ORF (or CDS) in 1250 distinct bacterial species, including 706 species whose genome had been fully sequenced and assembled at the time of analysis. This search essentially covered the ~3600 *Rho* entries that were present in InterPro (accession no. IPR004665; <http://www.ebi.ac.uk/interpro>). About 99.9% of these entries derive from *rho* sequences found in bacterial genomes, often from several strains of the same species (not shown). For instance, we retrieved more than 250 entries for *ecRho* corresponding to various *E. coli* strains or genome annotation projects. To reduce representation biases, analysis was thus conducted with a curated dataset that contained *Rho* sequence(s) from the most advanced annotation project for only one representative strain per bacterial species (Table S1, available in *Microbiology Online*).

Species containing *rho* belong to nearly all bacterial phyla with the noticeable exceptions of Tenericutes and Cyanobacteria (see below), which confirms previous observations (Opperman & Richardson, 1994; Washburn *et al.*, 2001). Analysis of published data also indicates that *Rho* proteins are expressed in representative species across phyla (Table S2). The putative (or experimentally detected) proteins encoded by the compiled *rho* ORFs have an average length of 480 aa, although the most frequent *Rho* length is closer to 420 aa (Fig. 2a). However, length distributions may be biased by disproportionate representations of phyla among the sequenced genomes. Indeed, the typical *Rho* length appears to vary significantly with the phylum and is noticeably larger for Actinobacteria and Bacteroidetes (Fig. 2b) due to the presence of NTD insertions and/or extensions (see below). Variations in *Rho* length are also larger within these phyla as well as among Firmicutes, Spirochaetes and members of the Planctomycetes–Verrucomicrobia–Chlamydia superphylum (Fig. 2b). We did not detect any correlation between *Rho* length and specific features of the corresponding genome such as its size, number of genes or GC content (data not shown).

The longest *Rho* sequence (865 aa) in our dataset is from the marine extremophile *Thermaerobacter marianensis* (Firmicutes). As observed for *Rho* proteins from other organisms, this is due to a long NTD region (~490 aa) upstream from the CSD-like RBD (Fig. 1b) that bears little

resemblance to other known sequences (no significant BLAST hit; data not shown). Although very short *Rho* sequences (<300 aa) are present in the dataset, they usually correspond to partial reads from uncompleted genome assemblies (WGS genomes in Table S1). In fact, the shortest *Rho* sequences deduced from fully assembled genomes are those of two γ -Proteobacteria, *Colwellia psychrerythraea* (314 aa) and *Marinomonas* sp. MWYL1 (318 aa), and both lack important RNA-binding motifs (RNP1 and RNP2; Fig. 1b) in their NTD. The two γ -Proteobacteria, however, contain another, longer *rho* ORF (Table S1) that encodes all key motifs supporting *Rho* activity, including canonical RNA-binding motifs (see below). Although it is unknown if *Rho* factors are expressed in these two bacteria (Table S2), these observations may indicate that one of the *rho* duplicates has been inactivated, possibly to prevent functional redundancy. Inactivation can also result from frame-shift mutations that disrupt *rho*, as is observed for *Dehalococcoides ethenogenes* (Chloroflexi) and *B. pumilus* (Firmicutes). In these cases, however, there is no evidence for a second *rho* ORF (Table S1), suggesting that *Rho* function is dispensable in these two micro-organisms.

We were surprised to detect *rho* in a few eukaryotic genomes. In two instances (*Drosophila ananassae* and *Callosobruchus chinensis*), *rho* presence appears to be due to horizontal gene transfers (HGTs) from *Wolbachia* endosymbionts (Dunning Hotopp *et al.*, 2007; Nikoh *et al.*, 2008). In other instances, however, we could not find a better explanation for the presence of *rho* than bacterial contamination of the genomic DNA used for sequencing.

The *rho* gene is absent from Cyanobacteria, Mollicutes and a fraction of Firmicutes

We did not find *rho* in any of the fully assembled genomes of Cyanobacteria (29 species) and Tenericutes (27 Mollicutes) that were present in GenBank at the time of our analysis (Table S3). We also did not find *rho* in the fully assembled genomes of 43 other bacterial species (Table S3), most of which are Firmicutes from Clostridia, Bacilli and Negativicutes classes. If one also considers uncompleted genome assemblies, it appears that a minority of Clostridia species but all negativicutes and most bacilli of the order Lactobacillales [e.g. *Streptococcus pneumoniae*, a species unaffected by BCM (Nishida *et al.*, 1972)] do not have *rho* (Tables S1 and S3; data not shown). By contrast, all bacilli of the order Bacillales (e.g. *B. subtilis* or *Staphylococcus aureus*) contain a *rho* CDS (Table S1). In recently constructed trees of Firmicutes (Wolf *et al.*, 2004; Zhao *et al.*, 2005; see also SILVA project at <http://www.arb-silva.de>), *rho*-less species are polyphyletic, suggesting that the *rho* gene was repeatedly lost in several, independent events during evolution.

Finally, we did not find *rho* in the fully assembled genomes of four Proteobacteria (Table S3). Three are insect endosymbionts that have very small genomes due to

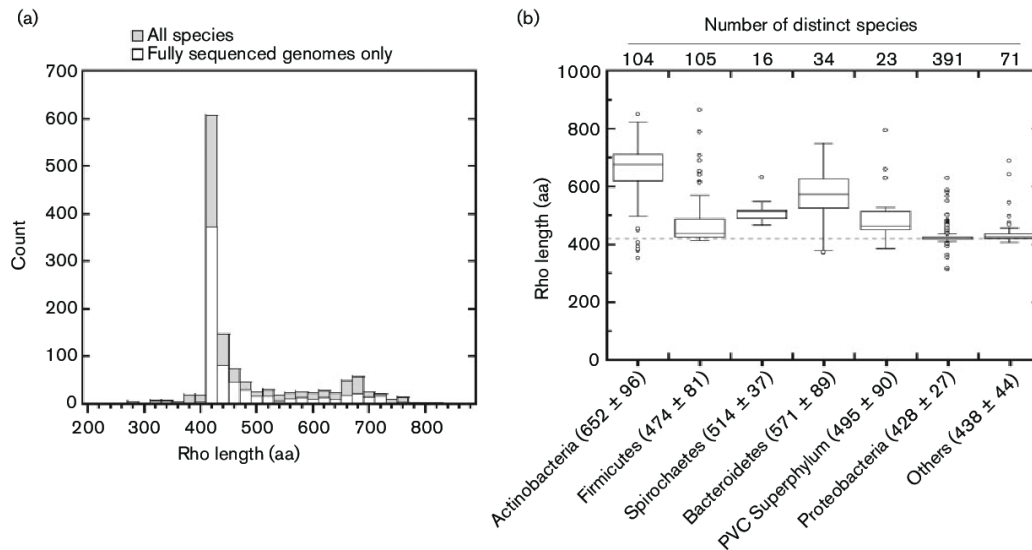


Fig. 2. Distribution of Rho lengths. (a) Histogram showing the distribution of the lengths of the putative Rho proteins corresponding to the complete set of identified locus tags (grey bars) or limited to fully sequenced genomes (white bars), as in Table S1. (b) Distribution of Rho lengths for representative phyla (fully sequenced genomes only). Boxplots show the median (inner line), interquartile range (IQR) [25th to 75th percentage (box)], acceptable range (whiskers) and outliers (open circles). The acceptable range is delineated by the lowest and highest data points that remain within a distance of $1.5 \times \text{IQR}$ from the box edges, as implemented in KALEIDAGRAPH software. Mean $\pm$ SD values are indicated in parentheses for each phylum. The dotted line corresponds to the most typical Rho length (420 aa).

reductive evolution processes (McCutcheon *et al.*, 2009). Massive gene loss and pseudogenization typically occur during adaptation to intracellular life (Bordenstein & Reznikoff, 2005), and it is possible that *rho* has been lost in the process. Reductive evolution may also explain the absence of *rho* in Mollicutes (Table S3), a class of bacteria with small genomes and parasitic behaviour (e.g. *Mycoplasma*). The obligate symbiotic/parasitic lifestyles of these organisms may have helped render Rho-dependent regulation superfluous. However, this determinant is not necessarily sufficient or compelling. For instance, obligate intracellular *Rickettsia* and *Chlamydia* contain CDS encoding canonical Rho proteins (Table S1). Furthermore, Rho is expressed in at least some *Chlamydia* species (Table S2).

We observed that *rho*-less species tend to have smaller, GC-poorer genomes than species containing *rho* ORFs ($P < 10^{-3}$; Fig. 3). This trend remains significant even when Mollicutes are excluded from the analysis (Fig. 3). The amount of GC is probably not an independent factor because it is itself positively correlated with genome size [for reasons that remain speculative (see Bentley & Parkhill, 2004; Hildebrand *et al.*, 2010)]. Our observations thus suggest that *rho* function is linked to some form of complexity in genome organization and/or expression, even though some exceptions are notable. For instance, we could not find traces of *rho* in the large, GC-rich genome

of *Haliangium ochraceum* (Table S3), a marine micro-organism (δ -Proteobacteria) with complex metabolic and morphogenesis features, including social motility and fruiting body behaviours (Ivanova *et al.*, 2010). The limited genomic information available for other members of the Nannocystineae suborder suggests that *rho* may be totally absent from this branch of Myxococcales (data not shown). This would be different from other myxobacteria, such as *Sorangium* or *Myxococcus* species, which bear at least one and often two *rho* ORFs within their large genomes (see Table S1). Considering that Rho is expressed in *Sorangium cellulosum* (Table S2), it will be interesting to determine if that is the case in other myxobacteria and why Rho is absent from members of the Nannocystineae suborder.

We did not detect other unusual features in genome architecture, such as specific biases in strand composition, or in orientation of CDSs, which would be characteristic of *rho*-less genomes (data not shown). We also did not detect obvious correlations between the absence of Rho and that of major transcription factors or Rho-interacting proteins. We found *hfq* genes and even a hypothetical *rof* (*yaeO*) gene in *rho*-less Firmicutes while we could not find *hfq* and/or *rof* genes in many Rho-containing species (data not shown). Moreover, we found transcription factors NusA and NusG in nearly all *rho*-less species, probably because these factors have other essential functions in transcription

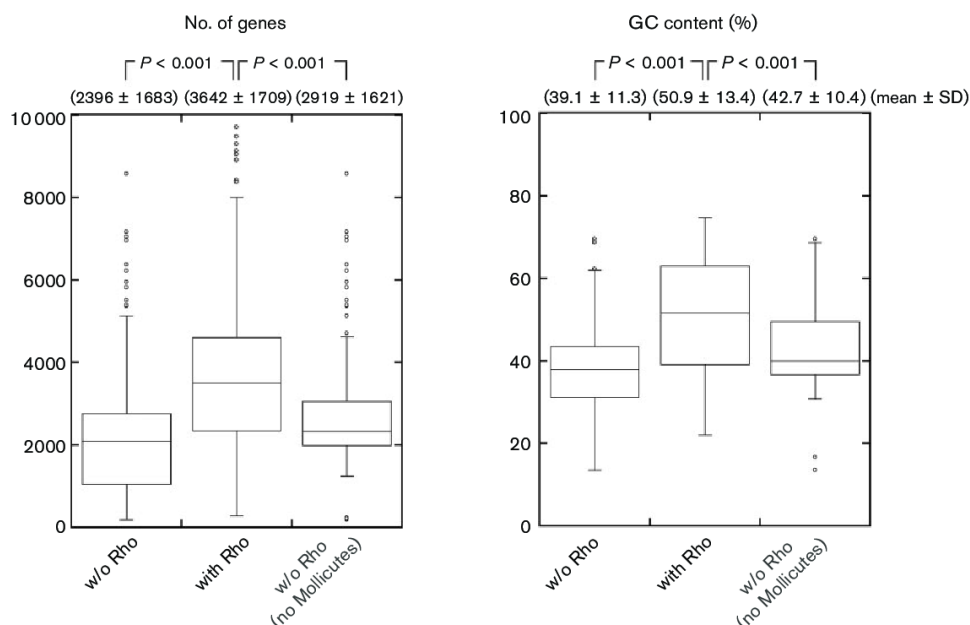


Fig. 3. Composition of the bacterial genomes surveyed. Only fully assembled genomes were considered (706 and 110 genomes with and without *rho*, respectively). See legend to Fig. 2b for definition of boxplots. *P*-values were determined with KALEIDAGRAPH and were similarly low with unpaired Wilcoxon–Mann–Whitney and *t*-test functions.

elongation and intrinsic termination mechanisms (Mooney *et al.*, 1998; Yang & Lewis, 2010). Intrinsic transcription termination mechanisms could actually compensate for the absence of Rho. In agreement with this proposal, computational searches uncovered high frequencies of putative intrinsic terminator motifs in *rho*-less genomes of Firmicutes (de Hoon *et al.*, 2005), Mollicutes (Kingsford *et al.*, 2007) and Cyanobacteria (Vijayan *et al.*, 2011). Notable exceptions are genomes of *Mycoplasma genitalium*, *My. pneumoniae* and *Synechocystis* PCC 6803 in which putative intrinsic terminators were only scarcely detected (de Hoon *et al.*, 2005; Kingsford *et al.*, 2007; Washio *et al.*, 1998). However, it remains currently uncertain if these exceptions are due to limitations of the computational search methods or if they reflect predominance of alternative, unidentified termination mechanisms (de Hoon *et al.*, 2005; Kingsford *et al.*, 2007).

Some bacterial genomes contain two *rho* ORFs

During our analysis, we observed that a number of species from diverse phyla contain two distinct *rho* ORFs (Table S1), a feature suggestive of HGT or chromosomal duplication. To better profile *rho* duplications, we have built a phylogenetic tree of all Rho duplicates found in the fully assembled bacterial genomes available at the time of analysis. As shown in Fig. 4, phylogenetic relationships

between Rho duplicates vary from close relatedness (e.g. in α -Proteobacteria) to deeply rooted divergence (e.g. in γ - and δ -Proteobacteria). Rho duplicates are monophyletic for α -Proteobacteria, Chloroflexi or Actinobacteria species (Fig. 4), suggesting that they originate from chromosomal duplication events. In the case of *Ehrlichia* and *Anaplasma* spp., synteny also supports chromosomal duplication/inversion events (Fig. S1), which tend to be frequent for these species (Dunning Hotopp *et al.*, 2006).

In contrast, γ - and δ -Proteobacteria, Planctomycetes or Bacteroidetes that contain two Rho homologues do not appear to have duplicated a pre-existing *rho* gene. Such organisms contain one Rho duplicate that clusters with Rho sequences from α -Proteobacteria (Fig. 4, group A), and a second that clusters with Rho sequences from Verrucomicrobia (group B). The great phylodivergence between duplicates from groups A and B (Fig. 4) and lack of obvious compositional bias or synteny (data not shown) make it difficult to assert the origin of these *rho* duplications without a more detailed analysis, which is beyond the scope of the present study.

As noted above, the absence of important RNA recognition motifs in group B's Rho factors from *C. psychrerythraea* and *Marinomonas* sp. MWYL1 (Fig. S2) should severely compromise their enzymic activity. Significant deletions within Rho's NTD are also apparent for additional

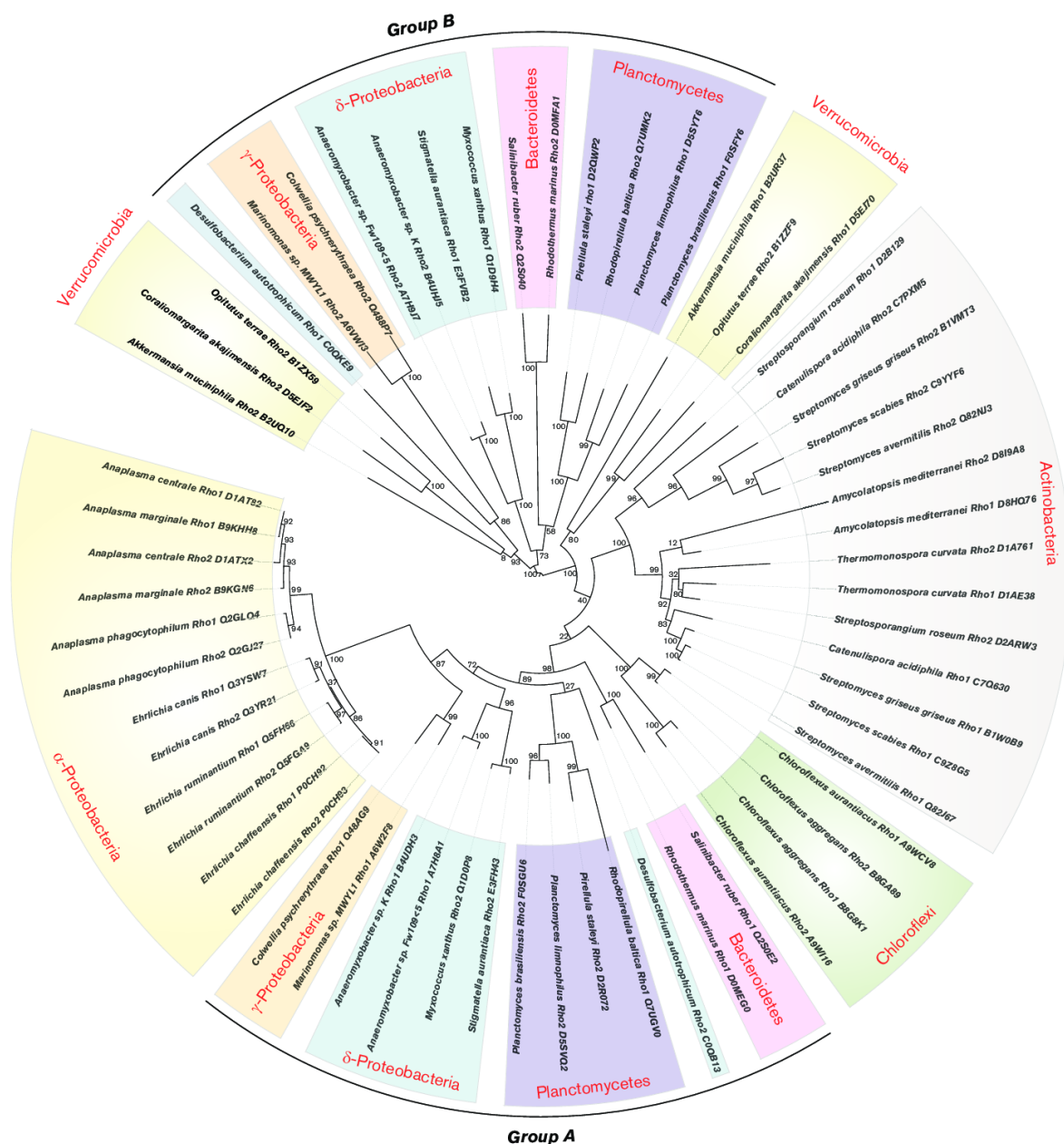


Fig. 4. Phylogram of the putative Rho proteins encoded by genomes containing *rho* duplicates. Numbers at nodes are bootstrap support values obtained upon maximum-likelihood analysis. Groups A and B contain Rho duplicates from γ - and δ -Proteobacteria, Planctomycetes and Bacteroidetes. For each species, one Rho duplicate belongs to group A, whereas the other falls within the phylo-distinct group B.

members of group B as well as for half of the Rho duplicates from Actinobacteria (Fig. 5, orange bars; data not shown). This may indicate that functional redundancy is often suppressed by enzymic inactivation of one of the

Rho duplicates. Alternatively, the altered Rho duplicate may have evolved towards a new function or mechanism that does not require the deleted domains, yet provides a selective advantage (Näsvall *et al.*, 2012). It is also possible

F. d'Heygère, M. Rabhi and M. Boudvillain

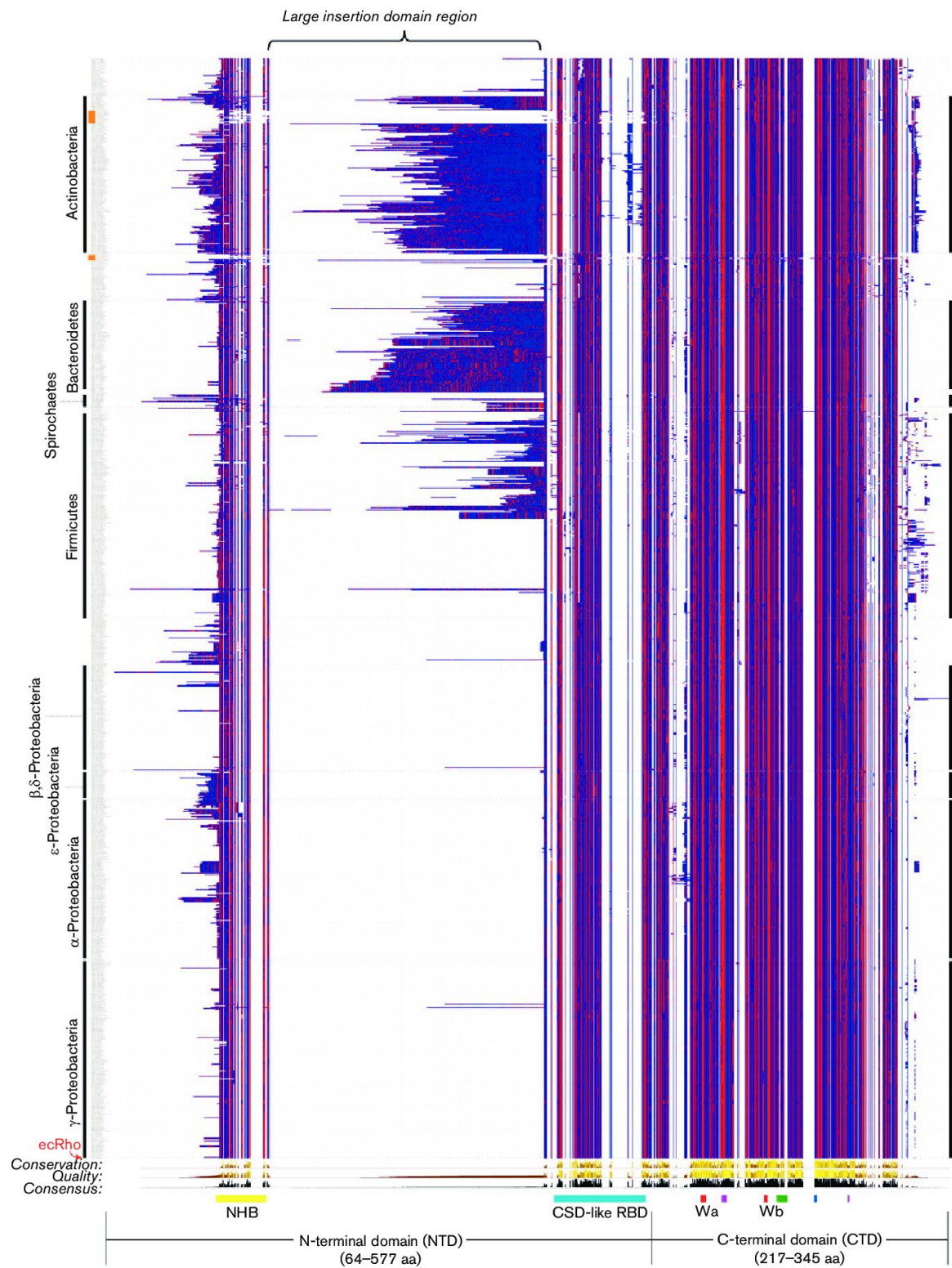


Fig. 5. Alignment overview of 1259 Rho sequences found across the bacterial kingdom (notable phyla are identified on the left side of the figure). The figure has been constructed with JALVIEW. Residues were colour-boxed using the hydrophobicity scale of JALVIEW (most hydrophobic and hydrophilic residues in red and blue, respectively). The huge number of sequences in the alignment overview required zooming out to a resolution that no longer allows identification of individual species (names on the left sides of sequences) or residues. For the same reason, the black dotted gaps in individual sequences appear as greyish zones in the figure. Histograms representing the levels of sequence conservation, quality of conservation and sequence consensus, as defined in JALVIEW (see Methods), are shown below the alignment. Coloured bars are also shown at the bottom to locate important motifs along the Rho sequence (colour coding as in Fig. 1c). Wa and Wb indicate Walker A and B motifs, respectively. Orange bars on the top left side of the figure identify unusually short Rho duplicates found in Actinobacteria and in group B of Fig. 4. The arrangement of phyla is arbitrary and not based on phylogenetic relatedness. The ecRho sequence is bottom last, as indicated on the left side of the figure.

that the Rho duplicate is not expressed at all. To conclude decisively, it will be necessary to determine the respective expression patterns of the Rho duplicates. This may be particularly relevant for species containing Rho duplicates that also bear aa substitutions conferring resistance to BCM (see below), an obvious selective advantage.

Sequence conservation in Rho homologues

To assess the conservation of important Rho motifs, we have aligned full-length Rho sequences listed in Table S1. Sequence alignments reveal a number of highly conserved regions along the Rho polypeptide that are well outlined by conservation histograms and by red/blue hydrophobic/phylic ‘vertical lines’ in Fig. 5 (due to the ‘zoomed out’ alignment of colour-boxed residues required to show all the 1259 Rho sequences). These regions include NHB and CSD-like RBD segments within Rho’s NTD, the region that connects the NTD and CTD, as well as most of Rho’s CTD (Fig. 5; see also below). Overall, Rho’s CTD appears more conserved than that of NTD, which confirms previous observations (Opperman & Richardson, 1994). Of particular notice is the fact that most Rho factors from Actinobacteria and Bacteroidetes, a significant fraction of Rho factors from Firmicutes and a few additional Rho factors from other phyla contain large insertion domains between the NHB and CSD-like RBD regions (Fig. 5). Altogether, these NTD insertion domains are found in ~35% of our compiled Rho sequences and, contrary to what was envisioned from limited datasets (Nowatzke *et al.*, 1997a), are not restricted to high G+C Gram-positive bacteria. Their sizes and compositions are not conserved (Fig. 5), although they often contain R/K-rich (mostly in Actinobacteria) or Q/N-rich (mostly in Bacteroidetes) stretches (data not shown). The NTD insertion domains are predominantly hydrophilic (Fig. 5 and data not shown; whole domain hydrophilic characters are also supported by negative GRAVY index values) but some are significantly basic (e.g. theoretical $pI > 9$ for the whole domain), while others are acidic (theoretical $pI < 6$). These insertion domains may promote specific Rho behaviours entailed by particular ecosystems, cellular metabolisms, genome structures and so on. For instance, the insertion domain of *Mc. luteus* Rho facilitates promoter-proximal transcription termination with structured RNA transcripts, a feature that

seems consistent with the high GC content of the *Mc. luteus* genome (Nowatzke *et al.*, 1997a). Similar observations have been made for the Rho factor from *Mb. tuberculosis* whose efficient termination activity in spite of poor ‘motor’ features may also be due to the presence of the insertion domain (Kalarickal *et al.*, 2010).

Overall, Actinobacteria contain the most atypical Rho factors with multiple insertions that alter the whole NTD (upstream from, between and within the NHB and CSD-like RBD) and that extend the CTD (Fig. 5). By contrast, Rho sequences from Proteobacteria form a rather homogeneous set of sequences, as evidenced by both length distribution (Fig. 2b) and alignment of sequences (Fig. 5). Homogeneity is particularly high when considering the γ -Proteobacteria alone (~235 species). Although a few Rho homologues contain NTD insertions, most γ -Proteobacteria sequences start right upstream from the NHB domain, as best exemplified by ecRho (Fig. S2 and data not shown). Given the importance of Rho in many γ -Proteobacteria, we have performed Shannon entropy analysis (Garcia-Boronat *et al.*, 2008) on the γ -Proteobacteria subset of Rho sequences (cleaned of NTD insertions) to better assess evolutionary sequence constraints imposed by Rho function. The γ -Proteobacteria-derived functional ‘core’ that was thereby analysed (Fig. 6, top) contains all the regions/motifs that are significantly conserved across phyla (Fig. 5). To facilitate interpretation of the entropy data, we have calculated a smoothed entropy curve whereby core motifs/regions of highest conservation are easily identified by local minima or ‘valleys’ (Fig. 6, red curve). For instance, the deepest entropy valley identifies the $\beta 7/\alpha 8$ junction region which contains the ‘catalytic’ Glu and Arg ‘valve’ residues (Fig. 6; Glu²¹¹ and Arg²¹² in ecRho) involved in ATP→ADP transitions (Thomsen & Berger, 2009) as the most conserved region of γ -Proteobacteria Rho proteins. This region is also extremely well conserved across phyla (see sequence logo in Fig. 6) with motif ERPEEVT (catalytic Glu and Arg valve residues in bold) found in all but two of the Rho sequences examined. Interestingly, the two divergent Rho sequences also contain non-conservative substitutions in other key motifs (Table S4 and data not shown) that suggest that Rho has been inactivated or relies on a different mechanism in the corresponding species (actinobacteria *Conexibacter woesei* and *Tropheryma whipplei*).

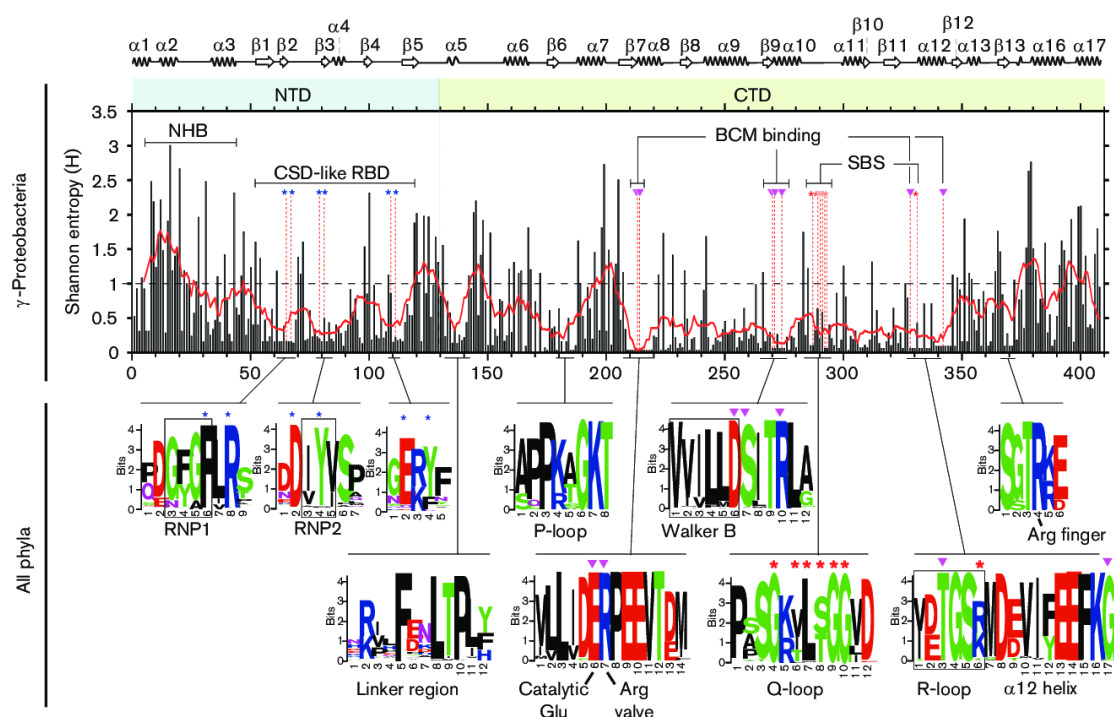


Fig. 6. Phyletic conservation of Rho. The histogram shows the Shannon entropy along the consensus sequence obtained upon alignment of Rho sequences from γ -Proteobacteria. Black bars show Shannon entropies calculated for the corresponding consensus positions, while the red curve represents the variation of local entropy after averaging over 10 surrounding positions (smoothed entropy curve). Blue and red stars indicate, respectively, the PBS and SBS residues identified in structures of *ec*Rho (PDB no. 1PVO and no. 3ICE). Pink triangles identify residues forming the BCM-binding pocket in *ec*Rho (PDB: no. 1XPO). The secondary structure diagram on top of the figure is based on PDB no. 1PVO. Sequence logos were prepared on the Weblogo server (<http://weblogo.berkeley.edu/>).

Three shallower entropy valleys identify the best conserved regions of Rho's NTD within the CSD-like RBD (Fig. 6). These regions include the RNP1 and RNP2 motifs as well as all the PBS residues that contact RNA/DNA ligands in crystal structures of *ec*Rho (Fig. 6, blue stars). These motifs and PBS residues are also relatively well conserved across phyla (Fig. 6; see also Table S4), although they are missing in a subset of Rho duplicates (Fig. 5, orange bars; data not shown). By contrast, the NHB subdomain is much less conserved among γ -Proteobacteria Rho factors or across phyla (Fig. 6 and Fig. S3). Hence, we have found only weak phylogenetic evidence to support the mechanism of electrostatic RNA recruitment by a positively charged NHB subdomain that has been inferred from the crystal structure of *Thermotoga maritima* Rho (Canals *et al.*, 2010).

Another region of high conservation in γ -Proteobacteria connects the NTD and CTD and contains a few residues that are also well conserved across phyla (Fig. 6). To the best of our knowledge, these residues have never been

subjected to systematic mutagenesis, although a K→Q substitution at position 130 (consensus position 132 in Fig. 6) spontaneously emerged in a genetic screen of impaired Rho function in *Salmonella* (Bossi *et al.*, 2012). Although it has been suggested that the linker region could be important for allosteric communication between the PBS and SBS (see Rabhi *et al.*, 2011a, and references within) or for interaction with NusG (Bossi *et al.*, 2012), these conjectures await experimental testing.

As is observed for the complete set of Rho sequences (Fig. 5), the most conserved region in γ -Proteobacteria is the central part of Rho's CTD (~210–350 region in Fig. 6). In addition to the $\beta 7/\alpha 8$ junction region discussed above, this segment of Rho's CTD includes the Walker B motif, Q- and R-loops, and the $\alpha 12$ helix, which have all been implicated in mechanochemical transduction within *ec*Rho (Thomsen & Berger, 2009). Key residues within these motifs, notably those contacting RNA in the central channel of *ec*Rho, are reasonably well conserved across phyla (Fig. 6, bottom), although a few Rho sequences

contain non-conservative substitutions (Table S4) that may be hallmarks of inactivation or alternative mechanochemistry (Adelman *et al.*, 2006). Side-chains forming the BCM-binding pocket in *ecRho* (Skordalakes *et al.*, 2005) are also highly conserved across phyla (Fig. 6, pink triangles) with only ~5 % of *rho*-containing species bearing a substitution at a critical position (Fig. S4). Interestingly, a majority of these substitutions are homologous to the single-point S266A and G337S substitutions that confer resistance to *E. coli* due to a weakened interaction between *ecRho* and BCM (Magyar *et al.*, 1999; Skordalakes *et al.*, 2005). This suggests that the corresponding bacterial species (Fig. S4) will also be poorly sensitive to BCM. Of particular notice are the cases of *Catenulispora* and *Streptomyces* species that contain Rho duplicates (Fig. S4 and Table S1). One of the duplicates almost systematically bears a substitution altering BCM binding (homologous to G337S; Fig. S4) as well as other changes in the CSD-like RBD (not shown). Whether these mutated Rho duplicates are expressed, functional and capable to confer resistance to BCM has not been determined. If true, this would contribute to explaining why *S. lividans*, which contains such a mutated Rho duplicate (Fig. S4), is poorly sensitive to BCM (Ingham *et al.*, 1996). Similarly, it will be worth checking if BCM-producing *Streptomyces* spp. also contain a mutated Rho duplicate and/or if they use alternative transcription termination (Kalarickal *et al.*, 2010) or efflux (Bentley *et al.*, 1993) mechanisms to outwit their own metabolite.

It is noteworthy that all of the Rho sequences from *Rickettsia* and fusobacteria species that are present in our dataset contain a substitution that confers a BCM-binding defect (Fig. S4 and Table S1). This suggests that all of these species, most of which are pathogens, will be poorly sensitive to BCM. It is also significant that substitutions conferring BCM-binding defects are found in both Gram-negative and Gram-positive bacteria (Fig. S4) from very diverse ecosystems (e.g. soils, mucus membranes, host cells). This is intriguing given that the only known natural sources for BCM are soil *Streptomyces* spp. (Miyamura *et al.*, 1973; Miyoshi *et al.*, 1972). In any case, the widespread occurrence of aa substitutions conferring resistance to BCM lends additional support to the view that Rho is important in a large and diverse set of bacterial species.

CONCLUSIONS

More than 90 % of the bacterial genomes that were sequenced at the time of our analysis contain a canonical *rho* gene (Tables S1 and S3). The prevalence of *rho* across the bacterial kingdom suggests that its origin is very ancient and that Rho-dependent functions are not easily substituted or discontinued in bacterial cells.

Rho is expressed in representative species across phyla (Table S2), yet is dispensable in some of them under

laboratory conditions (Quirk *et al.*, 1993; Washburn *et al.*, 2001). Even spurious transcription elicited by *rho* inactivation is not necessarily catastrophic, as observed for *B. subtilis* (Nicolas *et al.*, 2012), although condition-dependent or long-term effects may not be readily detected. For instance, genome instability may arise from uncontrolled encounters between transcription and replication machineries, since Rho inactivation should favour their co-directional collisions and the resultant formation of DNA double-strand breaks (Dutta *et al.*, 2011). Traffic control may also explain why Rho is sometimes more important or abundant under conditions such as cold (Mykytczuk *et al.*, 2011; Piette *et al.*, 2010; Quirk *et al.*, 1993) that favour formation (and thus require clearance) of roadblocks such as arrested RNA polymerases and R-loops (Erie, 2002; Massé & Drolet, 1999). Termination mechanisms that protect genomes from transcription-associated instability seem universal (Mischo *et al.*, 2011; Skourti-Stathaki *et al.*, 2011) and may thus well constitute the least common denominator of Rho function across the bacterial kingdom.

Small genomes tend to belong to species restricted to a stable niche, often with a parasitic/symbiotic lifestyle, whereas bacteria with large genomes tend to occupy highly complex and variable environments (Bentley & Parkhill, 2004). Since species without *rho* tend to have smaller genomes than species that contain *rho* (Fig. 3), requirement for *rho* may be linked, at least in part, to the complexity of the bacterium ecosystem and to the need for mechanisms that ensure protection/adaptation to environmental changes. Selective pressure for *rho* may notably arise from ‘immunity-like’ mechanisms whereby Rho contributes to repressing harmful transcription of HGT sequences (Cardinale *et al.*, 2008; Mitra & Nagaraja, 2012; Peters *et al.*, 2009). By limiting exposure to foreign DNA (Bordenstein & Reznikoff, 2005), the restrictive lifestyles of obligate intracellular bacteria may thus reduce the selective pressure for Rho function(s).

Although alterations of functionally important Rho motifs are sometimes observed, key residues involved in Rho’s oligomerization, ATP hydrolysis, RNA binding and RNA translocation are highly conserved (Fig. 6). The most noticeable disparity among Rho sequences is the presence of a large NTD insertion domain in about a third of them (Fig. 5). This insertion domain has been proposed to facilitate Rho binding to GC-rich transcripts in Actinobacteria (Nowatzke *et al.*, 1997a). However, its length and composition are not conserved and it is also found in Rho factors from GC-poor Firmicutes and Bacteroidetes (Fig. 5). Thus, the role(s) of the insertion domains remain(s) unclear and will have to be investigated for a large and representative set of species. Notably, it will be important to determine if insertion domains can elicit alternative termination mechanisms such as the one proposed for *Mb. tuberculosis* Rho (Kalarickal *et al.*, 2010). Together with the present study, this information should expand our understanding of Rho functions and mechanisms across the bacterial kingdom and may also

facilitate the rational design of high-throughput screening assays in the search for new antibiotics targeting Rho.

ACKNOWLEDGEMENTS

We gratefully acknowledge Drs Elena Bidnenko, Soraya Aït-Bara, Lionello Bossi and Nara Figueroa-Bossi for insightful discussions and critical reading of the manuscript. This work was supported by a joint PhD scholarship from the University of Orléans and Région Centre to F.D'H. and by CNRS core funding to M. B.

REFERENCES

- Adelman, J. L., Jeong, Y. J., Liao, J. C., Patel, G., Kim, D. E., Oster, G. & Patel, S. S. (2006). Mechanochemistry of transcription termination factor Rho. *Mol Cell* 22, 611–621.
- Banerjee, S., Chalisery, J., Bandey, I. & Sen, R. (2006). Rho-dependent transcription termination: more questions than answers. *J Microbiol* 44, 11–22.
- Bentley, S. D. & Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annu Rev Genet* 38, 771–792.
- Bentley, J., Hyatt, L. S., Ainley, K., Parish, J. H., Herbert, R. B. & White, G. R. (1993). Cloning and sequence analysis of an *Escherichia coli* gene conferring bicyclomycin resistance. *Gene* 127, 117–120.
- Bordenstein, S. R. & Reznikoff, W. S. (2005). Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* 3, 688–699.
- Bossi, L., Schwartz, A., Guillemardet, B., Boudvillain, M. & Figueroa-Bossi, N. (2012). A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes Dev* 26, 1864–1873.
- Boudvillain, M., Nollmann, M. & Margeat, E. (2010). Keeping up to speed with the transcription termination factor Rho motor. *Transcription* 1, 70–75.
- Boudvillain, M., Figueroa-Bossi, N. & Bossi, L. (2013). Terminator still moving forward: expanding roles for Rho factor. *Curr Opin Microbiol* 16, 118–124.
- Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B. & other authors (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433, 531–537.
- Canals, A., Usón, I. & Coll, M. (2010). The structure of RNA-free Rho termination factor indicates a dynamic mechanism of transcript capture. *J Mol Biol* 400, 16–23.
- Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. & Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* 320, 935–938.
- Ciampi, M. S. (2006). Rho-dependent terminators and transcription termination. *Microbiology* 152, 2515–2528.
- de Hoon, M. J., Makita, Y., Nakai, K. & Miyano, S. (2005). Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Comput Biol* 1, e25.
- Dunning Hotopp, J. C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S. V., Eisen, J. A., Seshadri, R., Ren, Q., Wu, M. & other authors (2006). Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* 2, e21.
- Dunning Hotopp, J. C., Clark, M. E., Oliveira, D. C., Foster, J. M., Fischer, P., Muñoz Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N. & other authors (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317, 1753–1756.
- Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* 146, 533–543.
- Erie, D. A. (2002). The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. *Biochim Biophys Acta* 1577, 224–239.
- Garcia-Boronat, M., Diez-Rivero, C. M., Reinherz, E. L. & Reche, P. A. (2008). PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res* 36 (Web Server issue), W35–W41.
- Gouy, M., Guindon, S. & Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27, 221–224.
- Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J. & Sassetti, C. M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* 7, e1002251.
- Hildebrand, F., Meyer, A. & Eyre-Walker, A. (2010). Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 6, e1001107.
- Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E. & Groisman, E. A. (2012). Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci U S A* 109, 5376–5381.
- Hu, P., Janga, S. C., Babu, M., Diaz-Mejia, J. J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S. & other authors (2009). Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol* 7, e96.
- Ingham, C. J., Hunter, I. S. & Smith, M. C. (1996). Isolation and sequencing of the rho gene from *Streptomyces lividans* ZX7 and characterization of the RNA-dependent NTPase activity of the overexpressed protein. *J Biol Chem* 271, 21803–21807.
- Ivanova, N., Daum, C., Lang, E., Abt, B., Kopitz, M., Saunders, E., Lapidus, A., Lucas, S., Glavina Del Rio, T. & other authors (2010). Complete genome sequence of *Haliangium ochraceum* type strain (SMP-2). *Stand Genomic Sci* 2, 96–106.
- Kalarickal, N. C., Ranjan, A., Kalyani, B. S., Wal, M. & Sen, R. (2010). A bacterial transcription terminator with inefficient molecular motor action but with a robust transcription termination function. *J Mol Biol* 395, 966–982.
- Kingsford, C. L., Ayanbule, K. & Salzberg, S. L. (2007). Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8, R22.
- Leela, J. K., Syeda, A. H., Anupama, K. & Gowrishankar, J. (2013). Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc Natl Acad Sci U S A* 110, 258–263.
- Li, J., Mason, S. W. & Greenblatt, J. (1993). Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes Dev* 7, 161–172.
- Magyar, A., Zhang, X., Abdi, F., Kohn, H. & Widger, W. R. (1999). Identifying the bicyclomycin binding domain through biochemical analysis of antibiotic-resistant rho proteins. *J Biol Chem* 274, 7316–7324.
- Massé, E. & Drolet, M. (1999). R-loop-dependent hypernegative supercoiling in *Escherichia coli* topA mutants preferentially occurs at low temperatures and correlates with growth inhibition. *J Mol Biol* 294, 321–332.
- McCutcheon, J. P., McDonald, B. R. & Moran, N. A. (2009). Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* 5, e1000565.

- Mischo, H. E., Gómez-González, B., Grzechnik, P., Rondón, A. G., Wei, W., Steinmetz, L., Aguilera, A. & Proudfoot, N. J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* 41, 21–32.
- Mitra, A. & Nagaraja, V. (2012). Under-representation of intrinsic terminators across bacterial genomic islands: Rho as a principal regulator of xenogenic DNA expression. *Gene* 508, 221–228.
- Miyamura, S., Ogasawara, N., Otsuka, H., Niwayama, S., Tanaka, H., Take, T., Uchiyama, T. & Ochiai, H. (1973). Antibiotic 5879 produced by *Streptomyces aizunensis*, identical with bicyclomycin. *J Antibiot (Tokyo)* 26, 479–484.
- Miyoshi, T., Miyairi, N., Aoki, H., Kosaka, M. & Sakai, H. (1972). Bicyclomycin, a new antibiotic. I. Taxonomy, isolation and characterization. *J Antibiot (Tokyo)* 25, 569–575.
- Mooney, R. A., Artsimovitch, I. & Landick, R. (1998). Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J Bacteriol* 180, 3265–3275.
- Myktyczuk, N. C., Trevors, J. T., Foote, S. J., Leduc, L. G., Ferroni, G. D. & Twine, S. M. (2011). Proteomic insights into cold adaptation of psychrotrophic and mesophilic *Acidithiobacillus ferrooxidans* strains. *Antonie van Leeuwenhoek* 100, 259–277.
- Näsvall, J., Sun, L., Roth, J. R. & Andersson, D. I. (2012). Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338, 384–387.
- Nicolas, P., Mäder, U., Dervyn, E., Roach, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M. & other authors (2012). Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335, 1103–1106.
- Nikoh, N., Tanaka, K., Shibata, F., Kondo, N., Hizume, M., Shimada, M. & Fukatsu, T. (2008). *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. *Genome Res* 18, 272–280.
- Nishida, M., Mine, Y., Matsubara, T., Goto, S. & Kuwahara, S. (1972). Bicyclomycin, a new antibiotic. 3. In vitro and in vivo antimicrobial activity. *J Antibiot (Tokyo)* 25, 582–593.
- Nowatzke, W. L., Burns, C. M. & Richardson, J. P. (1997a). Function of the novel subdomain in the RNA binding domain of transcription termination factor Rho from *Micrococcus luteus*. *J Biol Chem* 272, 2207–2211.
- Nowatzke, W. L., Keller, E., Koch, G. & Richardson, J. P. (1997b). Transcription termination factor Rho is essential for *Micrococcus luteus*. *J Bacteriol* 179, 5238–5240.
- Opperman, T. & Richardson, J. P. (1994). Phylogenetic analysis of sequences from diverse bacteria with homology to the *Escherichia coli* rho gene. *J Bacteriol* 176, 5033–5043.
- Peters, J. M., Mooney, R. A., Kuan, P. F., Rowland, J. L., Keles, S. & Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci U S A* 106, 15406–15411.
- Peters, J. M., Vangeloff, A. D. & Landick, R. (2011). Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol* 412, 793–813.
- Peters, J. M., Mooney, R. A., Grass, J. A., Jessen, E. D., Tran, F. & Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev* 26, 2621–2633.
- Pichoff, S., Alibaud, L., Guédant, A., Castanié, M. P. & Bouché, J. P. (1998). An *Escherichia coli* gene (yaeO) suppresses temperature-sensitive mutations in essential genes by modulating Rho-dependent transcription termination. *Mol Microbiol* 29, 859–869.
- Piette, F., D'Amico, S., Struvay, C., Mazzucchelli, G., Renaut, J., Tutino, M. L., Danchin, A., Leprince, P. & Feller, G. (2010). Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Mol Microbiol* 76, 120–132.
- Quirk, P. G., Dunkley, E. A., Jr, Lee, P. & Krulwich, T. A. (1993). Identification of a putative *Bacillus subtilis* rho gene. *J Bacteriol* 175, 8053.
- Rabhi, M., Tuma, R. & Boudvillain, M. (2010a). RNA remodeling by hexameric RNA helicases. *RNA Biol* 7, 655–666.
- Rabhi, M., Rahmouni, A. R. & Boudvillain, M. (2010b). Transcription termination factor Rho: a ring-shaped RNA helicase from bacteria. In *RNA Helicases*, vol. 19, pp. 243–271. Edited by E. Jankowsky. Cambridge, UK: RSC Publishing.
- Rabhi, M., Gocheva, V., Jacquino, F., Lee, A., Margeat, E. & Boudvillain, M. (2011a). Mutagenesis-based evidence for an asymmetric configuration of the ring-shaped transcription termination factor Rho. *J Mol Biol* 405, 497–518.
- Rabhi, M., Espéli, O., Schwartz, A., Cayrol, B., Rahmouni, A. R., Arluison, V. & Boudvillain, M. (2011b). The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J* 30, 2805–2816.
- Schmidt, M. C. & Chamberlin, M. J. (1984). Binding of rho factor to *Escherichia coli* RNA polymerase mediated by NusA protein. *J Biol Chem* 259, 15000–15002.
- Schwartz, A., Rabhi, M., Jacquino, F., Margeat, E., Rahmouni, A. R. & Boudvillain, M. (2009). A stepwise 2'-hydroxyl activation mechanism for the bacterial transcription termination factor Rho helicase. *Nat Struct Mol Biol* 16, 1309–1316.
- Skordalakes, E., Brogan, A. P., Park, B. S., Kohn, H. & Berger, J. M. (2005). Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure* 13, 99–109.
- Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* 42, 794–805.
- Thomsen, N. D. & Berger, J. M. (2009). Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* 139, 523–534.
- Tuckerman, J. R., Gonzalez, G. & Gilles-Gonzalez, M. A. (2011). Cyclic di-GMP activation of polynucleotide phosphorylase signal-dependent RNA processing. *J Mol Biol* 407, 633–639.
- Vijayan, V., Jain, I. H. & O'Shea, E. K. (2011). A high resolution map of a cyanobacterial transcriptome. *Genome Biol* 12, R47.
- Washburn, R. S. & Gottesman, M. E. (2011). Transcription termination maintains chromosome integrity. *Proc Natl Acad Sci U S A* 108, 792–797.
- Washburn, R. S., Marra, A., Bryant, A. P., Rosenberg, M. & Gentry, D. R. (2001). rho is not essential for viability or virulence in *Staphylococcus aureus*. *Antimicrob Agents Chemother* 45, 1099–1103.
- Washio, T., Sasayama, J. & Tomita, M. (1998). Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res* 26, 5456–5463.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. (2009). Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
- Williams, R. M. & Durham, C. A. (1988). Bicyclomycin: synthetic, mechanistic, and biological studies. *Chem Rev* 88, 511–540.
- Wolf, M., Müller, T., Dandekar, T. & Pollack, J. D. (2004). Phylogeny of Firmicutes with special reference to Mycoplasma (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. *Int J Syst Evol Microbiol* 54, 871–875.

F. d'Heygère, M. Rabhi and M. Boudvillain

Yang, X. & Lewis, P. J. (2010). The interaction between RNA polymerase and the elongation factor NusA. *RNA Biol* **7**, 272–275.

Zhao, Y., Davis, R. E. & Lee, I. M. (2005). Phylogenetic positions of 'Candidatus Phytoplasma asteris' and *Spiroplasma kunkelii* as inferred

from multiple sets of concatenated core housekeeping proteins. *Int J Syst Evol Microbiol* **55**, 2131–2141.

Edited by: R. Lan

Supplementary Tables

Table S3: experimental detection of Rho proteins in various bacterial species				
Phylum	Species	Method^a	Celular localisation^b	References
α -Proteobacteria	<i>Anaplasma phagocytophilum</i> ^c	LC-MS	n.d.	[1]
	<i>Bartonella quintana</i>	2D-PAGE	M	[2]
	<i>Brucella abortus</i>	2D-PAGE	C	[3]
	<i>Brucella melitensis</i>	2D-PAGE	n.d.	[4]
	<i>Candidatus Pelagibacter ubique</i>	LC-MS	n.d.	[5]
	<i>Rhodobacter capsulatus</i>	WB	n.d.	[6]
β -Proteobacteria	<i>Ralstonia eutropha</i>	2D-PAGE	SF	[7]
γ -Proteobacteria	<i>Acidithiobacillus ferrooxidans</i>	2D-PAGE	n.d.	[8]
	<i>Aeromonas hydrophila</i>	WB	n.d.	[9]
	<i>Enterobacter cloacae</i>	WB	n.d.	[9]
	<i>Erwinia chrysanthemi</i>	WB	n.d.	[9]
	<i>Escherichia coli</i>	WB	n.d.	[9]
		2D-PAGE	M	[10]
		2D-PAGE	C	[11]
		LC-MS	E>N	[12]
	<i>Haemophilus ducreyi</i>	2D-PAGE	C	[13]
	<i>Klebsiella pneumoniae</i>	WB	n.d.	[9]
	<i>Pasteurella multocida</i>	WB	n.d.	[9]
		LC-MS	n.d.	[14]
	<i>Proteus mirabilis</i>	WB	n.d.	[9]
	<i>Pseudoaltermonas haloplanktis</i>	2D-PAGE	n.d.	[15]
	<i>Pseudomonas aeruginosa</i>	WB	n.d.	[9]
		LC-MS	N	[12]
	<i>Pseudomonas putida</i>	2D-PAGE	n.d.	[16]
	<i>Salmonella typhimurium</i>	WB	n.d.	[9]
		LC-MS	C>E>OM	[17]
	<i>Serratia marcescens</i>	WB	n.d.	[9]
	<i>Shigella sonnei</i>	WB	n.d.	[9]
δ -Proteobacteria	<i>Sorangium cellulosum</i>	LC-MS	n.d.	[18]
ϵ -Proteobacteria	<i>Helicobacter mustelae</i>	LC-MS	C	[19]
	<i>Helicobacter pylori</i>	LC-MS	n.d.	[20]
Planctomycetes	<i>Rhodopirellula baltica</i>	2D-PAGE	n.d.	[21]
Chlamydiae	<i>Chlamydia trachomatis</i>	2D-PAGE	n.d.	[22]
	<i>Protochlamydia amoebophila</i>	2D-PAGE	n.d.	[23]
Bacteroidetes	<i>Porphyromonas gingivalis</i>	2D-PAGE	n.d.	[24]
	<i>Tannerella forsythia</i>	LC-MS	n.d.	[25]
Actinobacteria	<i>Mycobacterium bovis</i>	LC-MS	n.d.	[26]
	<i>Mycobacterium smegmatis</i>	LC-MS	M	[27]
	<i>Mycobacterium tuberculosis</i>	LC-MS	M	[28]
		LC-MS	M	[29]
		LC-MS	M	[30]
	<i>Mycobacterium ulcerans</i>	LC-MS	C	Database ^d
	<i>Streptomyces coelicolor</i>	LC-MS	n.d.	[31]
Firmicutes	<i>Bacillus subtilis</i>	WB	n.d.	[32]
		LC-MS	C	[33]
		LC-MS	E	[12]
	<i>Staphylococcus aureus</i>	1D-PAGE	M	[34]
		LC-MS	C	[35]

^a: LC, liquid chromatography; MS, mass spectrometry; WB, western blotting; PAGE, polyacrylamide gel electrophoresis.

^b: C, cytoplasm; E, envelope; OM, outer membrane; N, nucleoid; M, membrane; SF, soluble fraction; n.d., not determined.

^c: Identification peptides were exact matches to Rho1 and Rho2 proteins and therefore could not distinguish between the two factors.

^d: <http://genolist.pasteur.fr/BuruList/>

REFERENCES FOR TABLE S3

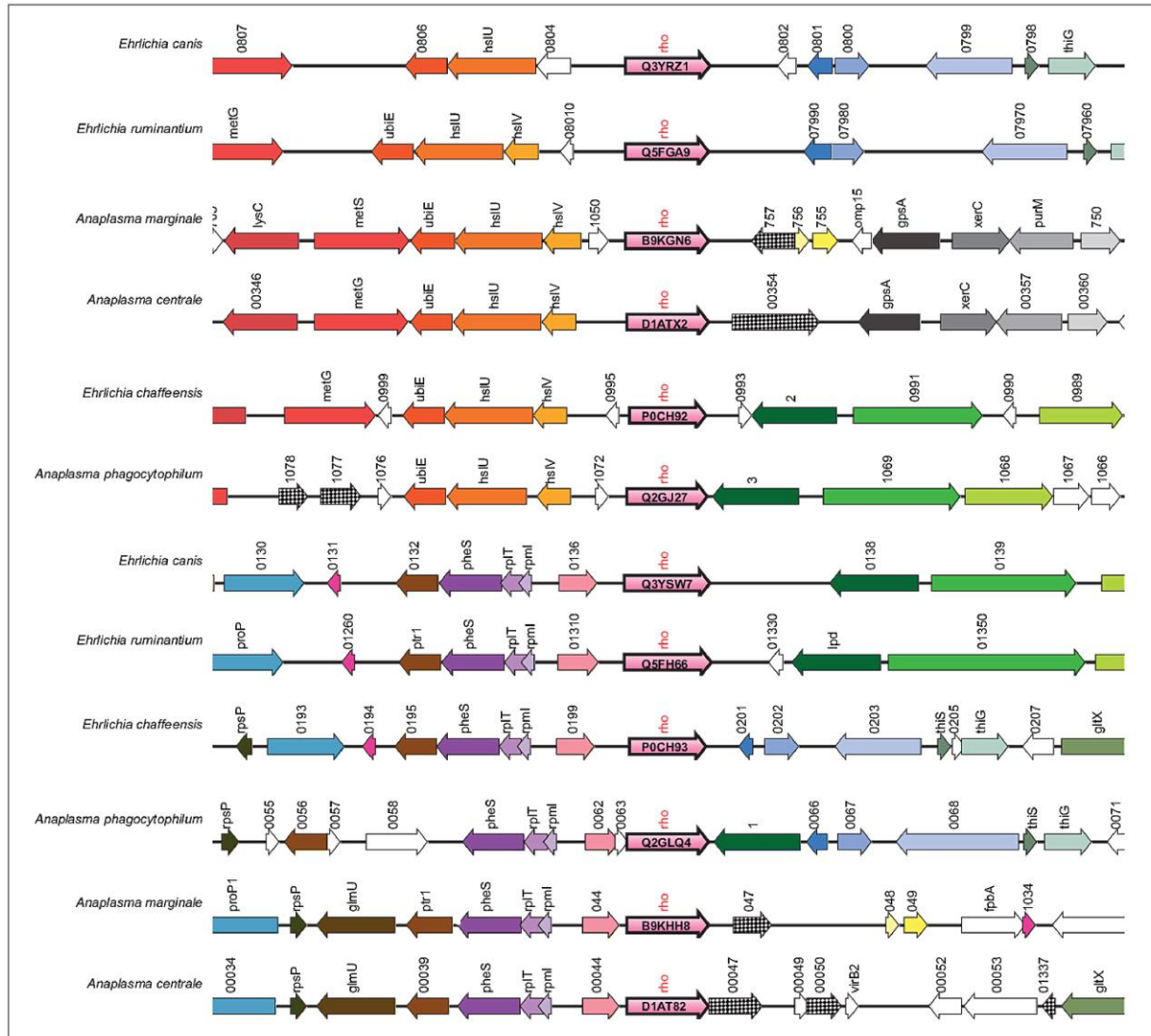
1. Troese, M.J., et al., *Infect Immun* (2011) 79:11, 4696-707.
2. Boonjakuakul, J.K., et al., *Infect Immun* (2007) 75:5, 2548-61.
3. Marchesini, M.I., J. Connolly, M.V. Delpino, P.C. Baldi, C.V. Mujer, V.G. DelVecchio, & D.J. Comerchi, *PLoS One* (2011) 6:12, e28480.
4. Wagner, M.A., et al., *Proteomics* (2002) 2:8, 1047-60.
5. Sowell, S.M., et al., *Appl Environ Microbiol* (2008) 74:13, 4091-100.
6. Jager, S., O. Fuhrmann, C. Heck, M. Hebermehl, E. Schiltz, R. Rauhut, & G. Klug, *Nucleic Acids Res* (2001) 29:22, 4581-8.
7. Schwartz, E., et al., *Proteomics* (2009) 9:22, 5132-42.
8. Myktyczuk, N.C., J.T. Trevors, S.J. Foote, L.G. Leduc, G.D. Ferroni, & S.M. Twine, *Antonie Van Leeuwenhoek* (2011) 100:2, 259-77.
9. Biville, F. & N. Guiso, *J Gen Microbiol* (1985) 131:11, 2953-60.
10. Fountoulakis, M. & R. Gasser, *Amino Acids* (2003) 24:1-2, 19-41.
11. Lopez-Campistrous, A., et al., *Mol Cell Proteomics* (2005) 4:8, 1205-9.
12. Ohniwa, R.L., Y. Ushijima, S. Saito, & K. Morikawa, *PLoS One* (2011) 6:4, e19172.
13. Scheffler, N.K., A.M. Falick, S.C. Hall, W.C. Ray, D.M. Post, R.S. Munson, Jr., & B.W. Gibson, *J Proteome Res* (2003) 2:5, 523-33.
14. Nanduri, B., M.L. Lawrence, D.S. Peddinti, & S.C. Burgess, *Comp Funct Genomics* (2008) 254836.
15. Piette, F., et al., *Mol Microbiol* (2010) 76:1, 120-32.
16. Sonawane, A., U. Kloppner, S. Hovel, U. Volker, & K.H. Rohm, *Microbiology* (2003) 149:Pt 10, 2909-18.
17. Coldham, N.G. & M.J. Woodward, *J Proteome Res* (2004) 3:3, 595-603.
18. Leinenbach, A., et al., *J Proteome Res* (2009) 8:9, 4350-61.
19. O'Toole, P.W., et al., *BMC Genomics* (2010) 11:164.
20. Jungblut, P.R., et al., *Proteomics* (2010) 10:2, 182-93.
21. Hieu, C.X., et al., *Proteomics* (2008) 8:8, 1608-23.
22. Saka, H.A., J.W. Thompson, Y.S. Chen, Y. Kumar, L.G. Dubois, M.A. Moseley, & R.H. Valdivia, *Mol Microbiol* (2011) 82:5, 1185-203.
23. Sixt, B.S., et al., *Proteomics* (2011) 11:10, 1868-92.
24. Zhang, Y., et al., *Proteomics* (2005) 5:1, 198-211.
25. Pham, T.K., S. Roy, J. Noirel, I. Douglas, P.C. Wright, & G.P. Stafford, *Proteomics* (2010) 10:17, 3130-41.
26. Jungblut, P.R., et al., *Mol Microbiol* (1999) 33:6, 1103-17.
27. He, Z. & J. De Buck, *BMC Microbiol* (2010) 10:121.
28. Gu, S., J. Chen, K.M. Dobos, E.M. Bradbury, J.T. Belisle, & X. Chen, *Mol Cell Proteomics* (2003) 2:12, 1284-96.
29. Mawuenyega, K.G., et al., *Mol Biol Cell* (2005) 16:1, 396-404.
30. de Souza, G.A., N.A. Leversen, H. Malen, & H.G. Wiker, *J Proteomics* (2011) 75:2, 502-10.
31. Kwangpyo, K., et al., *J Microbiol Biotechnol.* (2006) 16:5, 799-803.
32. Ingham, C.J., J. Dennis, & P.A. Furneaux, *Mol Microbiol* (1999) 31:2, 651-63.
33. Wolff, S., et al., *Mol Cell Proteomics* (2006) 5:7, 1183-92.
34. Scherl, A., et al., *J Microbiol Methods* (2005) 60:2, 247-57.
35. Kohler, C., et al., *Int J Med Microbiol* (2005) 295:8, 547-65.

Table S4: Bacterial species with deletions or non-conservative amino acid substitutions at critical Rho positions^a

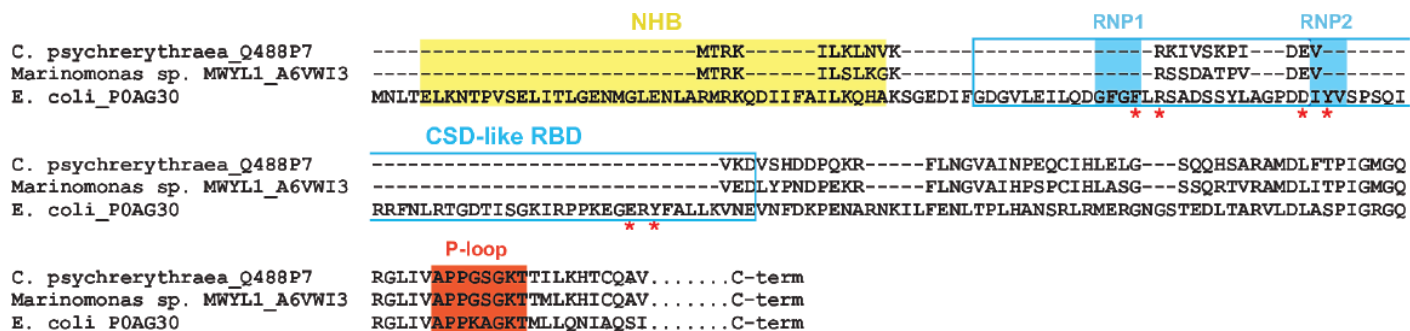
		RNP1		RNP2		cGlu	ICR	SBS	ICR	SBS			ICR		SBS	ICR						fArg	ICR
		F64	R66	D78	Y80	E211	N275	G282	K283	T286	G287	G288	D290	T323	K326	Y332	E333	K336	R347	E351	R353	R366	E368
Actinobacteria	Amycolatopsis mediterranei_Rho2_D8I9A8				X																		
	Actinobacteria subset (~70 species)													X									
	Conexibacter woesei_D3F2E5					X		X		X	X		X	X								X	X
	Streptomyces avermitilis_Rho2_Q82NJ3				X															X			
	Streptomyces flavogriseus_Rho2_E8VZR3	X	X		X																		
	Streptomyces ghanaensis_Rho2_D5ZY90	X			X															X			
	S. griseoaurantiacus_Rho2_F3NT71	X			X															X			
	S. griseus griseus_Rho2_B1VMT3	X			X																		
	Streptomyces lividans_Rho2_D6EDL5				X															X			
	Streptomyces scabies_Rho2_C9YVF6	X			X																		
	Streptomyces svicensis_B5I3Z7	X			X															X			
	S. viridochromogenes_Rho2_D9X8Z8	X			X															X			
	Streptomyces_sp._AA4_D9VFY7			X																X			
	Streptosporangium roseum_Rho1_D2B129				X					X													
	Thermomonospora curvata_Rho2_D1A761									X			X										
	Tropheryma whipplei_Q83MV6					X	X	X		X	X	X				X	X	X					
Deinococcus-thermus	Deinococcus deserti_C1CV53		X		X																		
	Deinococcus geothermalis_Q1IZD3		X		X																		
	Deinococcus radiodurans_P52153		X		X																		
	Truepera radiovictrix_D7CSK5		X		X																		
	Oceanithermus profundus_E4U8C2		X		X																		
	Thermus aquaticus_B7A688		X	X	X																		
	Thermus thermophilus_Q72JR8		X	X	X																		
	Meiothermus ruber_D3PRP1		X		X																		
Firmic.	Meiothermus silvanus_D7BH57		X		X																		
	Pseudoramibacter alactolyticus_E6MEV0																			X			
	Gemella haemolysans_C5NWQ7																	X	X				
Planctomyc.	Gemella moribillum_E5V2H5																	X	X				
	Planctomyces brasiliensis_Rho1_F0SFY6															X							
	Planctomyces limnophilus_Rho1_D5SYT6															X							
	Pirellula staleyi_rho1_D2QWP2							X								X							
	Rhodopirellula baltica_Rho1_Q7UGV0																		X				
δ-Prot.	Rhodopirellula baltica_Rho2_Q7UMK2															X							
	Anaeromyxobacter sp._K_Rho2_B4UHI5				X																		
	A. sp._Fw109-5_Rho2_A7H9J7				X																		
	D. autotrophicum_Rho1_C0QKE9															X							
	Myxococcus xanthus_Rho1_Q1D9H4				X																		
γ-Prot.	Stigmatella aurantiaca_Rho1_E3FVB2				X																		
	Enhydrobacter aerosaccus_C8Q0N3													X									
	Ruthia magnifica_A1AWC6															X							
Verrucomicrobia	Vesicomysocius okutanii_A5CWY2															X							
	Akkermansia muciniphila_Rho1_B2UQ10			X												X				X			
	Akkermansia muciniphila_Rho2_B2UR37	X																					
	Coraliomargarita akajimensis_Rho1_D5EJ70	X	X																				
	Coraliomargarita akajimensis_Rho2_D5EJF2	X	X													X							
	Opitutus terrae_Rho1_B1ZX59	X	X													X							
	Opitutus terrae_Rho2_B1ZZF9		X																				

^a: The identities of critical residues, as defined from studies of *ecRho*, are given at the top of the table. ICR, intersubunit contacting residue; SBS, secondary binding site; cGlu, catalytic glutamate; fArg, arginine 'finger' residue.

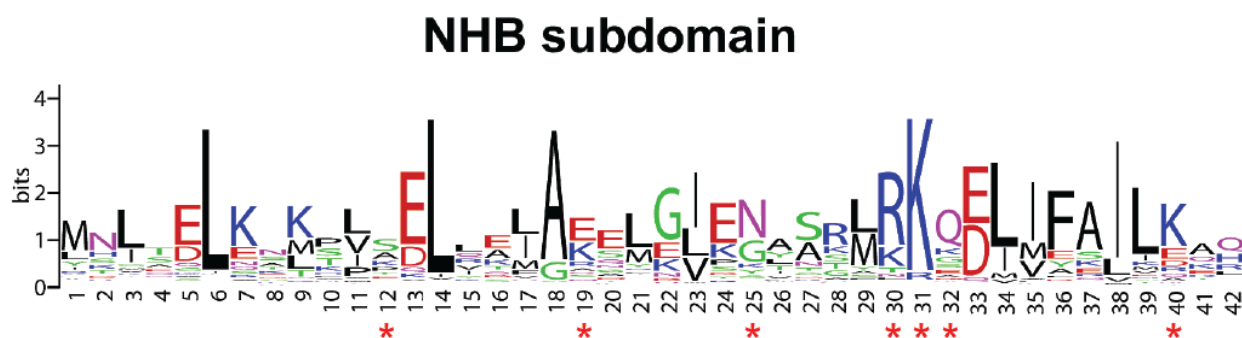
Supplementary Figures



Supplementary Fig. S1: Synteny between chromosomal regions containing *rho* duplicates in α -proteobacteria. Genomic comparisons were performed on the Absynte server (<http://archaea.u-psud.fr/absynte>). White and shaded arrows correspond to unmatched genes and non-coding sequences, respectively.



Supplementary Fig. S2: Sequence alignment of the N-terminal regions of group B-Rho factors from the γ -proteobacteria *C. psychrerythraea* and *Marinomonas* sp. MWYL1. The sequence of the Rho factor from *E. coli* is also shown for comparison. Red stars locate PBS residues that contact RNA in *ecRho*.



Supplementary Fig. S3: Sequence logo showing the conservation across all phyla of the most relevant residues of the NHB subdomain (positions of potential amino acid insertions are not indicated). The positions of the basic residues that create a positively charged patch in the NHB subdomain of *T. maritima* Rho are identified by red stars.

Fusobacteria	<i>Fusobacterium_sp._3_1_33_D0BRQ8</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._7_1_C3WWK4</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._D11_D6BDP9</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_nucleatum_nucleatum_Q8RG42</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._4_1_13_C3WRT1</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._7_1_C3WWK4</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_nucleatum_vincentii_Q7P4H6</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._3_1_27_D6LCN6</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._3_1_36A2_C7XML1</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_periodonticum_D4CUZ0</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._1_1_41FAA_D6LGU8</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._2_1_31_C3WHJ8</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_varium_C6JQ77</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_mortiferum_C3W9S2</i>	E	R	D	S	R	T	S
	<i>Fusobacterium_sp._3_1_5R_E5BG66</i>	E	R	D	S	R	T	S
Actinobacteria	<i>Fusobacterium_sp._D12_E5BIV7</i>	E	R	D	S	R	T	S
	<i>Tropheryma_whipplei_Q83MV6</i>	I	R	D	S	H	S	A
	<i>Acidimicrobium_ferrooxidans_C7M177</i>	E	R	D	G	R	T	G
	<i>Catenulisporea_acidiphila_Rho2_C7PXM5</i>	E	R	D	S	R	T	S
	<i>Streptomyces_viridochromogenes_Rho2_D9X8Z8</i>	E	R	D	S	R	T	S
	<i>Streptomyces_lividans_Rho2_D6EDL5</i>	E	R	D	S	R	T	S
	<i>Streptomyces_svaceus_Rho2_B5I3Z7</i>	E	R	D	S	R	T	S
	<i>Streptomyces_scabies_Rho2_C9YYF6</i>	E	R	D	S	R	T	S
	<i>Streptomyces_flavogriseus_Rho2_E8VZR3</i>	E	R	D	S	R	T	S
	<i>Streptomyces_avermitilis_Rho2_Q82NJ3</i>	E	R	D	S	R	T	S
	<i>Streptomyces_griseoaurantiacus_Rho2_F3NT71</i>	E	R	D	S	R	T	S
	<i>Conexibacter_woesei_D3F2E5</i>	A	R	D	T	-	L	G
	<i>Amycolatopsis_mediterranei_D8I9A8</i>	E	R	D	S	R	T	S
	<i>Streptomyces_sp._AA4_D9VIFY7</i>	E	R	D	S	R	T	S
	<i>Nocardioideis_sp._BAA-499_A1SHH5</i>	E	R	D	G	R	S	G
Chloroflexi	<i>Aeromicrobium_marinum_E2SCP3</i>	E	R	D	G	R	T	G
	<i>Propionibacterium_acnes_D4HE63</i>	E	R	D	G	R	T	G
	<i>Propionibacterium_sp._409-HC1_F5TPU6</i>	E	R	D	G	R	T	G
	<i>Propionibacterium_humerusii_F3P177</i>	E	R	D	G	R	T	G
	<i>Propionibacterium_sp._434-HC2_F5TVB9</i>	E	R	D	G	R	T	G
	<i>Propionibacterium_freudenreichii_D7GDE6</i>	E	R	D	G	R	T	G
	<i>Dehalococcoides_sp._CBDB1_Q3ZYU3</i>	E	R	D	G	R	T	G
	<i>Dehalogenimonas_lykanthroporepellens_D8K298</i>	E	R	D	G	R	T	G
Spirochaetes	<i>Borrelia_hermsii_B2RZT9</i>	E	R	D	S	R	T	S
	<i>Borrelia_duttonii_B5RL54</i>	E	R	D	S	R	T	S
	<i>Borrelia_recurrentis_B5RR47</i>	E	R	D	S	R	T	S
	<i>Borrelia_turicatae_A1QZ23</i>	E	R	D	S	R	T	S
	<i>Borrelia_burgdorferi_P33561</i>	E	R	D	S	R	T	S
	<i>Borrelia_sp._SV1_C0APE2</i>	E	R	D	S	R	T	S
	<i>Borrelia_spielmanii_C0APP8</i>	E	R	D	S	R	T	S
	<i>Borrelia_garinii_Q662D4</i>	E	R	D	S	R	T	S
	<i>Borrelia_valaisiana_D6RX74</i>	E	R	D	S	R	T	S
	<i>Borrelia_afzelii_Q0SNT1</i>	E	R	D	S	R	T	S
Chrysiogenetes	<i>Desulfurispirillum_indicum_E6W467</i>	E	R	D	S	R	T	S
	<i>Wolbachia_pipientis_wMel_Q73GY4</i>	E	R	D	N	R	T	G
	<i>Wolbachia_endosymbiont_Drosophila_ananassae_Q4EAR7</i>	E	R	D	N	R	T	G
	<i>Wolbachia_sp._Drosophila_simulans_C0R3B3</i>	E	R	D	N	R	T	G
α -Proteobacteria	<i>Rickettsia_endosymbiont_Ixodes_scapularis_C4YUU1</i>	E	R	D	A	R	T	G
	<i>Rickettsia_prowazekii_Q9ZD24</i>	E	R	D	A	R	T	G
	<i>Rickettsia_bellii_Q1RIJ6</i>	E	R	D	A	R	T	G
	<i>Rickettsia_akari_A8GNQ7</i>	E	R	D	A	R	T	G
	<i>Rickettsia_africae_C3PNN5</i>	E	R	D	A	R	T	G
	<i>Rickettsia_conorii_Q92HL2</i>	E	R	D	A	R	T	G
	<i>Rickettsia_massilliae_A8F1S4</i>	E	R	D	A	R	T	G
	<i>Rickettsia_peacockii_C4K0C7</i>	E	R	D	A	R	T	G
	<i>Rickettsia_rickettsii_A8GSD0</i>	E	R	D	A	R	T	G
	<i>Rickettsia_sibirica_Q7P8M3</i>	E	R	D	A	R	T	G
	<i>Rickettsia_canadensis_A8EYQ8</i>	E	R	D	A	R	T	G
	<i>Rickettsia_felis_Q4ULF7</i>	E	R	D	A	R	T	G
	<i>Rickettsia_typhi_Q68WL0</i>	E	R	D	A	R	T	G
	<i>Escherichia_coli_P0AG30</i>	E	R	D	S	R	T	G
γ -Proteobacteria								

211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

Supplementary Fig. S4: Species with mutated residues in the BCM-binding pocket. *E. coli*'s residues are shown for comparison (numbers refer to residue positions in ecRho). Mutations already known to confer BCM-resistance to ecRho are in bold red.

ARTICLE II:
ATP-DEPENDENT MOTOR ACTIVITY OF
THE TRANSCRIPTION TERMINATION FACTOR RHO FROM
MYCOBACTERIUM TUBERCULOSIS

Nucleic Acids Research (2015)

Published online 20 May 2015

Nucleic Acids Research, 2015, Vol. 43, No. 12 6099–6111
doi: 10.1093/nar/gkv505

ATP-dependent motor activity of the transcription termination factor Rho from *Mycobacterium tuberculosis*

François D’Heygère^{1,2}, Annie Schwartz¹, Franck Coste¹, Bertrand Castaing^{1,3} and Marc Boudvillain^{1,3,*}¹Centre de Biophysique Moléculaire, CNRS UPR4301, rue Charles Sadron, 45071 Orléans cedex 2, France, ²Ecole doctorale Santé, Sciences Biologiques et Chimie du Vivant (ED 549), Université d’Orléans, Orléans, France and ³ITP Sciences Biologiques & Chimie du Vivant, Université d’Orléans, Orléans, France

Received March 23, 2015; Revised April 28, 2015; Accepted May 04, 2015

ABSTRACT

The bacterial transcription termination factor Rho—a ring-shaped molecular motor displaying directional, ATP-dependent RNA helicase/translocase activity—is an interesting therapeutic target. Recently, Rho from *Mycobacterium tuberculosis* (M_{tb} Rho) has been proposed to operate by a mechanism uncoupled from molecular motor action, suggesting that the manner used by Rho to dissociate transcriptional complexes is not conserved throughout the bacterial kingdom. Here, however, we demonstrate that M_{tb} Rho is a *bona fide* molecular motor and directional helicase which requires a catalytic site competent for ATP hydrolysis to disrupt RNA duplexes or transcription elongation complexes. Moreover, we show that idiosyncratic features of the M_{tb} Rho enzyme are conferred by a large, hydrophilic insertion in its N-terminal ‘RNA binding’ domain and by a non-canonical R-loop residue in its C-terminal ‘motor’ domain. We also show that the ‘motor’ domain of M_{tb} Rho has a low apparent affinity for the Rho inhibitor bicyclomycin, thereby contributing to explain why *M. tuberculosis* is resistant to this drug. Overall, our findings support that, in spite of adjustments of the Rho motor to specific traits of its hosting bacterium, the basic principles of Rho action are conserved across species and could thus constitute pertinent screening criteria in high-throughput searches of new Rho inhibitors.

INTRODUCTION

RNA helicases are ubiquitous motor proteins that remodel RNA and RNA-protein structures and are involved in all steps of RNA metabolism (1,2). Their activities are often

implicated in infection and disease and thus constitute attractive prognostic/diagnostic markers or potential drug targets (3). RNA helicases generally rely on one of two distinct mechanisms of action (1,4). Some RNA helicases are directional, NTP-fueled translocase enzymes exhibiting some degree of processivity. In this case, the helicase loads onto a single-stranded RNA segment and then uses the energy derived from ATP hydrolysis to migrate toward one end of the RNA chain, removing bound obstacles (such as a complementary strand) from its path. This mechanism is reminiscent of that used by DNA helicases (5). Other RNA helicases are local ‘disruptases’ that load on RNA and remodel/disrupt its local structure (or local interaction with an RNA-binding protein) in an ATP-dependent fashion. In this case, the RNA helicase acts as a non-directional and non-processive molecular switch. Local ‘disruptases’ are best exemplified by DEAD-box proteins which compose the largest group of RNA helicases. Upon ATP binding, DEAD-box proteins adopt a tight RNA-bound, ‘close’ state able to denature local RNA structures; upon release of the products of ATP hydrolysis, they revert to an RNA-free ‘open’ state allowing enzyme recycling (1,6).

The bacterial transcription termination factor Rho is an important model of ring-shaped helicase (4,5). Rho unwinds deleterious RNA:DNA duplexes (R-loops) that are formed behind RNA polymerase (RNAP) during transcription and dissociates transcription elongation complexes (TECs) at specific loci of bacterial genomes (7,8). The molecular mechanisms used by Rho to unwind duplexes and dissociate TECs have been studied extensively, although work has focused primarily on the Rho factor from *Escherichia coli* (E_c Rho) (7,9). These studies revealed that Rho uses a sophisticated translocation mechanism, called ‘tethered tracking’, to slide in a 5–3’ direction along nascent RNA, dissociating transcriptional R-loops and/or TECs in the process (Figure 1A). This mechanism invokes net translocation steps of one RNA base per hydrolyzed

*To whom correspondence should be addressed. Tel: +33 2 38 25 55 85; Fax: +33 2 38 63 15 17; Email: marc.boudvillain@cnrs-orleans.fr

© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

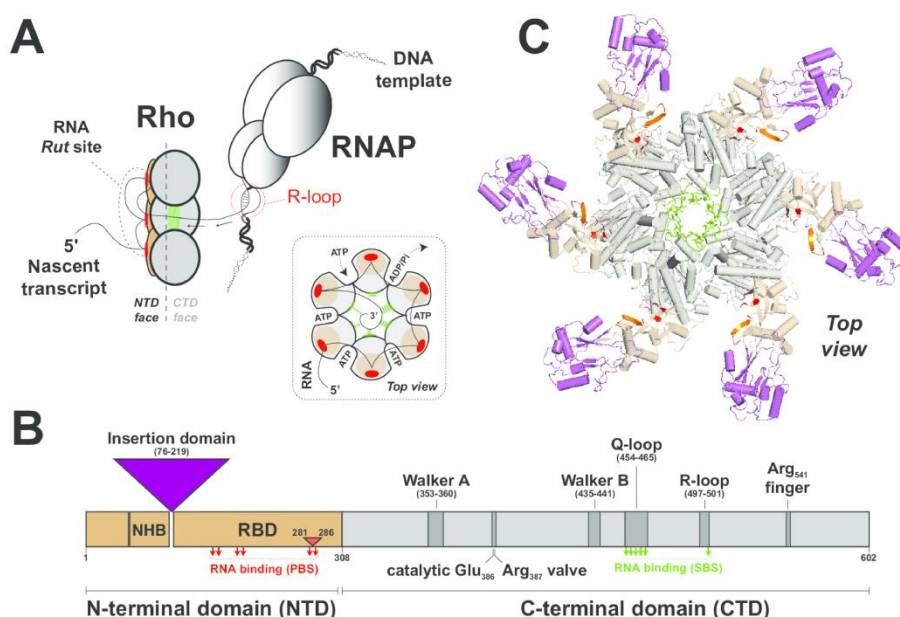


Figure 1. Transcription termination factor Rho. (A) A model of Rho-dependent transcription termination based on studies of *E. coli* Rho (7). Once bound to a C-rich *Rut* site, the Rho hexamer translocates along the nascent transcript in an ATP-dependent fashion until it catches up with RNAP and triggers dissociation of the TEC. Obstacles such as transcriptional R-loops are removed in the process. Direct contacts between Rho and RNAP, which may exist prior to Rho binding to the nascent transcript (55), are not represented here. Similarly, the exact pathway leading to TEC dissociation once Rho has caught up with the RNAP is still debated (7,8) and is not detailed here. A schematic of the RNA-bound Rho hexamer during ‘tethered tracking’ is also shown in the inset. (B) Diagram showing the arrangement of key motifs within the primary sequence of Rho subunits. The N-terminal helix bundle (NHB) domain may contribute to RNA recruitment to the Rho hexamer (36) which is ensured primarily by residues within the RNA binding domain (RBD). The large insertion domain present in *M. tuberculosis* Rho (13) is depicted by a purple triangle. Amino acid numbering corresponds to the *M. tuberculosis* Rho protein. PBS and SBS residues contacting RNA in *E. coli* Rho structures (10,37) are depicted by red and green arrows, respectively. The 6 aa insert between PBS residues of *M. tuberculosis* Rho is depicted by a small orange triangle. (C) A model of the *M. tuberculosis* Rho hexamer. An homology model of the *M. tuberculosis* Rho subunit was built with Swiss-Model (<http://www.swissmodel.expasy.org/>) without pre-defined template. Then, the *M. tuberculosis* Rho hexamer was reconstructed by aligning modeled *M. tuberculosis* Rho subunits with subunits of the *E. coli* Rho hexamer (PDB: 3ICE) using Pymol software (<http://www.pymol.org>). Apart from the insertion domain, the *M. tuberculosis* Rho and *E. coli* Rho subunits are structurally similar (rmsd ~ 1.3 Å). Color coding is the same as in panel B.

ATP molecule (10–12) and conforms to the general idea that toroidal helicases directionally translocate nucleic acids within their central channel (5).

Rho is an interesting antibacterial target because it is widespread in bacteria and absent from eukaryotes (13). Bicyclomycin (BCM), a natural Rho inhibitor used in veterinary medicine (14), is mostly effective against Gram-negative bacteria (15). Until now, only one Gram-positive species—*Micrococcus luteus*—has been found to be sensitive to BCM (16). Bacterial resistance to BCM may have several origins. Rho may be dispensable in (as for *Staphylococcus aureus*) (17) or absent from (as for ~8% of bacterial species) (13) a given bacterium. Alternatively, the bacterium may use a specific resistance mechanism, for instance one conferred by a drug efflux protein (18) or by mutations in Rho that alter BCM binding (13,19). It is also conceivable that the mechanism of action of Rho is different and insensitive to BCM in some bacteria. BCM binds in a subsite of the ATPase pocket of *E. coli* Rho where it occludes binding of the nucleophilic water molecule required for ATP hydrolysis (15,20). Thus, Rho factors that would not use ATPase-driven, molecular motor activity to trigger transcription termination would be inherently resistant to BCM.

Rho is considered essential for the growth of *Mycobacterium tuberculosis* (21–23), a species that is nonetheless resistant to BCM (24). A recent report suggests that *M. tuberculosis* Rho (*M. tuberculosis* Rho) has no duplex unwinding (helicase) activity and can dissociate TECs without efficient molecular motor action (25). The authors argue that only ATP binding to *M. tuberculosis* Rho is required for its transcription termination function because non-hydrolysable adenosine 5'-(β , γ -imido)triphosphate (AMP-PNP) and ATP γ S analogs stimulate the dissociation of artificially stalled TECs by *M. tuberculosis* Rho. They also observed that BCM inhibits the ATPase activity of *M. tuberculosis* Rho (IC_{50} = 30 μ M) but fails to significantly affect transcription termination in the tested concentration range (up to 200 μ M of BCM) (25). These unexpected findings suggest that *M. tuberculosis* Rho uses a mechanism that is different from that of *E. coli* Rho and reminiscent of that used by RNA helicases working as local ‘disruptases’.

To test the possibility that *M. tuberculosis* Rho uses a molecular switch (e.g. local ‘disruptase’) mechanism that would make it intrinsically resistant to BCM, we have studied the motor activity of *M. tuberculosis* Rho in detail using *in vitro* ATPase, duplex unwinding and transcription termination assays as well as a catalytic *M. tuberculosis* Rho mutant and *M. tuberculosis* Rho/*E. coli* Rho do-

main fusion chimera. We show that M_{tb} Rho is a [5'→3']-directional, ATP-dependent RNA helicase which requires a catalytic site that is fully competent for ATP hydrolysis to promote duplex unwinding or transcription termination. Non-hydrolysable ATP analogs do not promote duplex unwinding by M_{tb} Rho nor change dramatically M_{tb} Rho affinity for RNA. Furthermore, both the helicase and transcription termination activities of M_{tb} Rho can be inhibited by BCM, albeit at concentrations that are significantly higher than for E_c Rho. Experiments with a domain deletion mutant and the M_{tb} Rho/ E_c Rho chimera indicate that this low apparent BCM affinity is conferred principally by the C-terminal domain (CTD) of M_{tb} Rho while the N-terminal domain (NTD) specifies which regions of nascent RNA transcripts are utilized by M_{tb} Rho. Overall, our data argue against a scenario whereby M_{tb} Rho would rely on a distinctive, motor-independent mechanism (25) and demonstrate that M_{tb} Rho is an ATP-dependent translocase that is well adapted to the peculiar transcriptome and slow metabolism of *M. tuberculosis*. Our findings also suggest that inhibition of the ATPase-dependent activities of M_{tb} Rho could be a valid strategy in search of new antibiotics targeting *M. tuberculosis*.

MATERIALS AND METHODS

Materials

Chemicals and enzymes were purchased from Sigma-Aldrich and New England Biolabs, respectively. BCM was purchased from Santa Cruz Biotechnology. Sigma-saturated RNAP from *E. coli* was obtained from Epicentre Biotechnologies. Nucleoside triphosphates were purchased from GE-Healthcare while ADP, AMP-PNP and ATPγS analogs were from Sigma-Aldrich. Radionucleotides were from PerkinElmer. Synthetic oligonucleotides were obtained from Eurogentec. Polynucleotide fragment (> 300 nt) stocks were prepared as described previously (26) from polynucleotide batches obtained from Santa Cruz Biotechnology and Midland Certified Reagent Company (Midland, TX, USA). Streptavidin-coated magnetic beads were from Ademtech (Pessac, France). DNA templates used in *in vitro* transcription reactions were prepared by standard PCR procedures, as described previously (27). RNA substrates were obtained by *in vitro* transcription of PCR amplicons with T7 RNAP and purified by denaturing polyacrylamide gel electrophoresis (PAGE), as described (26). Plasmids for over-expression of single-point M_{tb} Rho mutants were prepared by Quickchange (Stratagene) mutagenesis of the pET28b- M_{tb} Rho plasmid (kindly provided by Dr. Ranjan Sen, Hyderabad, India). Plasmids for over-expression of the ΔM_{tb} Rho deletion mutant (wherein amino acids 76–219 of M_{tb} Rho are replaced by a Val-Pro dipeptide) and the $[M_{tb}:E_c]$ Rho (wherein amino acids 1–308 of M_{tb} Rho are fused to amino acids 131–419 of E_c Rho) and $[E_c:M_{tb}]$ Rho (wherein amino acids 1–130 of E_c Rho are fused to amino acids 309–602 of M_{tb} Rho) chimeras were obtained by inserting synthetic DNA fragments (obtained by overhang extension PCR using the pET28b-Rho (26) and/or pET28b- M_{tb} Rho plasmids as starting templates) between the NdeI and XhoI sites of the pET28b plasmid (Novagen) following

standard cloning procedures (see Supplementary Table S1 for details). The sequences of DNA templates and plasmids were verified by capillary DNA sequencing (Genoscreen, France).

Preparation of proteins

The E_c Rho protein was prepared and purified as described previously (26). Wild-type (WT) and mutant M_{tb} Rho factors as well as M_{tb} : E_c protein chimeras were over-expressed as N-terminal His tag fusions (25) in Rosetta 2(DE3) cells (Merck-Millipore) harboring the appropriate pET28b derivative and were purified following published procedures (28). Briefly, the M_{tb} Rho proteins were affinity selected on His-Pur colbalt resin (Thermo scientific) and were subsequently purified by cation exchange chromatography on a SP sepharose FF column (GE-Healthcare) with a 0.15–1M NaCl gradient and by gel filtration chromatography on a Sephacryl S-300 HR column (GE-Healthcare) equilibrated in GF buffer (150 mM NaCl, 20 mM HEPES, pH 7.5, 2 mM β -mercaptoethanol). Protein preparations were controlled by liquid chromatography high-resolution mass spectrometry using an UltiMate 3000 NanoRSLC system (Dionex, Sunnyvale, CA, USA) connected to a 4-GHz MaXis ultra-high resolution quadrupole-TOF spectrometer (Bruker Daltonics) equipped with an electrospray ion source. All proteins were stored at -20°C as micromolar solutions in 100 mM KCl, 10 mM Tris-HCl, pH 7.9, 0.1 mM EDTA, 0.1 mM DTT and 50% (v/v) glycerol. Concentrations of E_c Rho and M_{tb} Rho proteins are expressed in hexamers throughout the manuscript. Purified Ded1 protein was kindly provided by Prof. Eckhard Jankowsky (Case Western Reserve University, Cleveland, USA).

Preparation of duplex substrates

Duplexes were assembled by mixing 10 pmoles of RNA transcript with 12 pmoles of complementary oligonucleotide (bearing a 6-carboxyfluorescein label at the 5'-end) in annealing buffer (150 mM potassium acetate, 20 mM HEPES, pH 7.5, 1 mM EDTA). Mixtures were heated at 95°C and slowly cooled to 20°C . Duplexes were then purified by native 7% PAGE and stored at -20°C in helicase buffer (20 mM HEPES, pH 7.5, 0.1 mM EDTA, 0.5 mM DTT and 150 mM potassium glutamate) (26).

NTP hydrolysis assay

The NTP hydrolysis activities of the Rho enzymes were determined with the EnzCheck Phosphate Assay kit (Molecular Probes) as described previously (27,28). Briefly, two quartz cuvettes containing equal volumes of a solution of 20 nM Rho, RNA cofactor (e.g. poly[rC]) at a standard concentration of 10 μM in rC residues, 0.4 mM 2-amino-6-mercapto-7-methylpurine riboside and 1 U/mL of purine nucleoside phosphorylase in NTPase buffer (1 mM MgCl_2 , 50 mM Tris-HCl, pH 7.5, 0.1 mM sodium azide) were placed in the control and sample beam holders of an Uvikon-XL UV-spectrophotometer equipped with a 37°C circulating water bath. After equilibration at 37°C for 5 min, equal volumes of water and NTP (1 mM, final concentration)

were added, respectively, to the control and reaction cuvettes and absorbance at 360 nm was recorded as a function of time. The NTPase rates were deduced from the initial first-derivative maxima that were obtained after smoothing of the time-dependent A_{360} curves with Kaleidagraph software. The experiments were calibrated with standard samples containing known amounts of phosphate, as recommended by the kit manufacturer.

Equilibrium binding assays

Equilibrium Rho-RNA dissociation constants were determined using a filter-binding assay, as described previously (29,30). Briefly, ~10 fmoles of 32 P-labeled RNA substrate were mixed with various amounts of Rho in 100 μ l of binding buffer (20 mM HEPES, pH 7.5, 0.1 mM EDTA, 0.5 mM DTT, 150 mM potassium acetate and 20 μ g/ml BSA) containing 0 or 1 mM ATP (or AMP-PNP). After incubation for 10 min at 30°C, the samples were filtered through stacked [top] nitrocellulose (Amersham Protran) and [bottom] cationic nylon (Pall Biotyne B) membranes using a Bio-dot SF apparatus (Biorad). No significant differences were observed in control experiments performed with 30 min incubation times (not shown). The fractions of free and Rho-bound RNA (retained on the nylon and nitrocellulose membranes, respectively) as a function of Rho concentration were then determined by phosphorimaging of the membranes using a Typhoon-Trio imager and dedicated ImageQuant TL v8.1 software (GE healthcare).

Duplex unwinding assay

Helicase reactions were performed as described previously (31) with 6-carboxyfluorescein labeled duplex substrates (control experiments with 32 P-labeled substrates supported that the fluorescein moiety does not affect M_{th} Rho behavior) (28). Briefly, 5 nM substrate was mixed with 50 or 500 nM Rho and 0–300 μ M BCM in helicase buffer and incubated for 3 min at 30°C. Then, 1 mM MgCl_2 , 1 mM nucleotide cofactor (ATP, ADP, $\text{ATP}\gamma\text{S}$, AMP-PNP, ADP- BeF_x or ADP- AlF_4) and 400 nM oligo trap (unlabeled oligonucleotide having the same sequence than the released ‘reporter’ strand) were added to the helicase mixture before further incubation at 30°C. Reaction aliquots were taken at various times and mixed with two volumes of quench buffer (10 mM EDTA, 1.5% SDS, 300 mM sodium acetate, 6% Ficoll-400) before being loaded on 9% polyacrylamide gels that contained 1X TBE and 0.5% SDS. Detection and quantification of gel bands were performed by fluorescence imaging with a Typhoon-Trio imager as described (28). Helicase reaction parameters were obtained by fitting data points to an equation describing the kinetic regimen uncovered previously for E_c Rho:

$$F_p = A \times (1 - e^{-k_{\text{exp}}t}) + k_{\text{lin}} \times t$$

where F_p is the fraction of product formed, A is the amplitude of the exponential (burst) phase of the reaction, and k_{exp} and k_{lin} are the rate constants of the exponential and linear phases of the reaction, respectively (30,32).

Transcription termination experiments

Transcription termination experiments were performed under single-round ‘chase’ conditions as described previously (33) with minor modifications. Briefly, biotinylated DNA templates were prepared by PCR amplification of pT7A1- λ cro DNA template (34) using a 5'-biotinylated reverse primer. To allow preparation of TECs halted at +24, a forward mutagenic PCR primer was also used to introduce a C→T mutation at position +6 of the pT7A1- λ cro DNA template. The resulting mAS97Bt template (30 nM) was mixed with *E. coli* RNAP holoenzyme (36 nM) in transcription initiation buffer (40 mM Tris-HCl, pH 7.9, 50 mM KCl, 5 mM MgCl_2 , 1 mM DTT, 0.05 mg/ml BSA, 0.4 U/ μ l Superase-In [Ambion]) and incubated at 37°C for 5 min. The mixture was then supplemented with 12 μ M ApU, 6 μ M ATP, 6 μ M GTP, 0.95 μ M UTP, and 1 μ Ci/ μ l [α - 32 P]-UTP and incubated for 10 min at 37°C. Then, the mixture was chilled on ice, KCl concentration was raised to 0.1 M and the halted transcription complexes were purified by filtration through a microspin G50 column pre-equilibrated with transcription elongation buffer (40 mM Tris-HCl, pH 7.9, 100 mM KCl, 5 mM MgCl_2 , 1 mM DTT, 0.1 mg/ml BSA, 0.2 U/ μ l Superase-In). The complexes were then immobilized on streptavidin beads (Supplementary Figure S4) or used directly in single-run chase reactions. In the latter case, an aliquot of the solution of halted transcription complexes (~15 nM) was supplemented with 70 nM Rho and 0–750 μ M BCM in transcription elongation buffer and incubated for 10 min at 37°C. Then, rNTPs (75 μ M each, final concentrations) and rifampicin (25 μ g/ml) were added to initiate chase reactions. After incubation for 10 min at 37°C, the mixtures were mixed with 1 volume of denaturing buffer (95% formamide, 50 mM EDTA) and resolved by 8–10% denaturing PAGE. Gels were analyzed by Typhoon-Trio phosphorimaging. Apparent global termination efficiencies (TE_{app}) were defined as:

$$\text{TE}_{\text{app}} = \frac{\sum I_{\text{term}}}{\sum I_{\text{term}} + \sum I_{\text{runoff}}} \times 100$$

where $\sum I_{\text{term}}$ is the sum of the intensities of the bands corresponding to termination products while $\sum I_{\text{runoff}}$ is the sum of the intensities of the bands corresponding to runoff products. IC_{50} values for BCM inhibition were obtained by non-linear least-square fitting of TE_{app} values obtained as a function of BCM concentration to a standard sigmoid inhibition equation:

$$\text{TE}_{\text{app}} = (\text{TE}_{\text{app}})_0 - F_{\text{max}} \times \frac{[\text{BCM}]^n}{[\text{BCM}]^n + [\text{IC}_{50}]^n}$$

where $(\text{TE}_{\text{app}})_0$ is the value of TE_{app} at 0 μ M BCM, F_{max} is the maximal fraction of TE_{app} that is sensitive to BCM inhibition and n is an empirical parameter that defines the sigmoid shape of inhibition (35).

Data reporting

All data values are reported as mean $\pm$ standard deviation calculated from at least two independent experimental replicates (whenever number of replicates $N_R > 2$ is indicated in figure legends).

RESULTS

Selection of protein mutants and chimeras to probe *Mtb* Rho features

The typical *Ec*Rho ring is made of six identical protein subunits, each divided in two distinct domains (7–9,36,37). The NTD contains primary binding site (PBS) motifs involved in the recruitment of RNA *Rut* (Rho utilization) sites by the Rho hexamer (Figure 1A and B). PBS motifs include residues forming a cleft only large enough to accommodate a 5'-YC dimer (Y and C being pyrimidine and cytosine residues, respectively), thereby explaining the preference of *Ec*Rho for single-stranded C-rich *Rut* sites (7–9,37). The CTD contains residues required for intersubunit cohesion and for 'molecular motor' function. Essential CTD motifs include Walker A and B motifs forming the ATP binding pockets at subunit interfaces, 'catalytic Glu', 'Arg valve' and 'Arg finger' residues required for the catalysis of ATP hydrolysis as well as Q-loop and R-loop residues forming the secondary binding site (SBS) for RNA (Figure 1B) (10,38). Importantly, changes in the chemical state of the ATPase pockets are allosterically transmitted up to the SBS residues which are in charge of RNA translocation within the central ring channel (Figure 1A) (10).

Alignment of Rho sequences from over 1200 bacterial species indicates that most of these key residues are conserved across the bacterial kingdom, including in phylo-distinct Rho factors such as *Mtb*Rho (Supplementary Figure S1) (13). Notwithstanding, *Mtb*Rho displays a few distinctive features that could explain its uncharacteristic behavior. These features include (i) a 144 amino acids (aa) insertion subdomain in the NTD (Figure 1B) that should result in bulky appendages to the hexameric ring (Figure 1C, in magenta); (ii) a short insert (aa 281–286) between PBS residues (Figure 1B and Supplementary Figure S1) that could affect the structure of the 5'-YC binding clefts (Figure 1C, in red) and recognition of *Rut* sites; and (iii) a threonine instead of a lysine at an SBS position in the R-loop (Thr₅₀₁ in *Mtb*Rho versus Lys₃₂₆ in *Ec*Rho; Supplementary Figure S1 and Figure 1C, in green) that could weaken the allosteric interaction network critical for Rho mechanochemistry.

To probe the effect of these distinctive features, we prepared *Mtb*Rho derivatives containing a single-point SBS mutation (T501A and T501K mutants) or lacking the NTD insertion subdomain (Δ *Mtb*Rho) and a protein chimera wherein the NTD of *Mtb*Rho is fused to the CTD of *Ec*Rho ([*Mtb:Ec*]Rho chimera). Multiple attempts to prepare the 'reverse' [*Ec:Mtb*]Rho chimera (NTD of *Ec*Rho fused to the CTD of *Mtb*Rho) were unsuccessful because this protein was invariably expressed in inclusion bodies and could not be purified in a soluble form under native conditions (data not shown).

To probe the role of ATP hydrolysis in *Mtb*Rho, we also prepared a single-point mutant wherein the catalytic glutamate residue Glu₃₈₆ is replaced by alanine (E386A mutant). The role of the catalytic glutamate residue is to polarize the nucleophilic water molecule for attack of the γ -phosphoryl group of ATP (10). The non-conservative mutation of this critical residue in *Ec*Rho (E211A mutation)

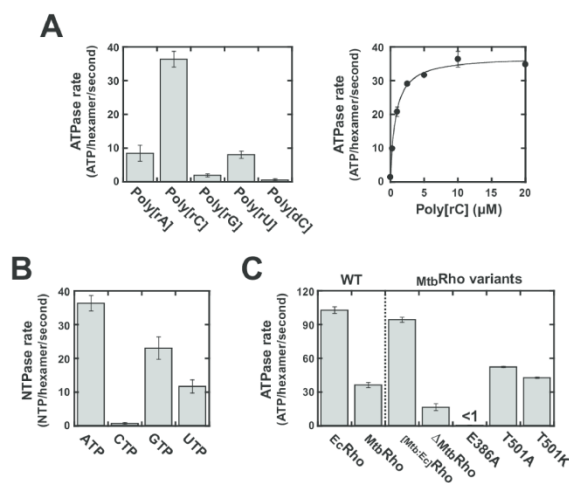


Figure 2. RNA-dependent NTP hydrolysis activities of the *Mtb*Rho protein variants. Unless specified otherwise, the concentrations of NTP and RNA polymer were 1 mM and 10 μ M (in nucleotide residues), respectively. (A) Stimulation of the ATPase activity of WT *Mtb*Rho by polyribonucleotide cofactors (number of replicates, $N_R = 4$ for 10 μ M poly[rC]). (B) WT *Mtb*Rho hydrolyzes ATP most efficiently in the presence of poly[rC] ($N_R = 4$ for ATP). (C) Comparison of the ATPase rate of WT *Mtb*Rho with rates measured under the same standard conditions for WT *Ec*Rho and *Mtb*Rho variants ($N_R = 3$ for Δ *Mtb*Rho and $N_R = 4$ for WT *Mtb*Rho).

does not affect ATP binding or protein oligomerization but completely abolishes ATP hydrolysis (38).

Tight allosteric coupling within the *Mtb*Rho enzyme

To assess the substrate/cofactor preferences of *Mtb*Rho, we performed standard measurements of its NTP hydrolysis rate (27) in the presence of various NTP substrates and polynucleotide cofactors. In agreement with previous reports (25,39), we found that an RNA cofactor is required to activate NTP hydrolysis by *Mtb*Rho (Figure 2A and data not shown). Poly[rC] is the preferred polynucleotide cofactor with a K_m of ~ 0.7 μ M (Figure 2A) while ATP is the preferred substrate (Figure 2B). Hydrolysis of CTP is highly inefficient (Figure 2B) at a concentration (1 mM) that is saturating for ATP ($K_{m, ATP} \sim 70$ μ M) (25), a feature that has also been observed for the Rho factors of *M. luteus* (*Mlut*Rho) (40) and *Streptomyces lividans* (41), two evolutionary relatives. By contrast, the phylodivergent *Ec*Rho displays little preference for a specific NTP (40).

Under our standard conditions (see 'Materials and Methods' section), the rate of ATPase turnover is about 3-fold lower for *Mtb*Rho than for *Ec*Rho or for the [*Mtb:Ec*]Rho chimera (Figure 2C), suggesting that components in the CTD of *Mtb*Rho are suboptimal. Deletion of the NTD insertion subdomain of *Mtb*Rho (Δ *Mtb*Rho) reduces the ATPase rate by half under standard conditions (Figure 2C). However, a higher concentration of poly[rC] cofactor is required to reach maximal ATPase turnover with the Δ *Mtb*Rho variant (data not shown), which is consistent with an RNA binding defect induced by the deletion of the NTD subdomain (see Table 1 and below). At saturating poly[rC]

concentrations, Δ_{Mtb} Rho hydrolyzes ATP at a rate of ~ 80 molecules per hexamer and per second, which is about twice faster than WT Mtb Rho (Figure 2C) and suggests that the NTD subdomain somewhat contributes to regulate ATP consumption.

Altering the identity of Thr₅₀₁ in Mtb Rho increases the rate of ATP hydrolysis by 1.2–1.4-fold (Figure 2C, T501A and T501K mutations). This modest effect suggests that Thr₅₀₁ has a less important role in Mtb Rho than has the corresponding Lys₃₂₆ side chain in the R-loop of Ec Rho (the K326A mutation reduces the ATPase activity of Ec Rho by $\sim 75\%$) (10,27). In sharp contrast, non-conservative mutation of the catalytic Glu₃₈₆ residue (E386A mutant) completely inhibits ATP hydrolysis by Mtb Rho, even when concentrations of ATP and poly[rC] are increased to 5 and 30 μM , respectively (Figure 2C and data not shown). This dramatic effect induced by the E386A mutation is not matched by a drastic change of the oligomeric state or capacity to bind ATP or RNA (Supplementary Figure S2 and Table 1). Thus, the E386A phenotype is comparable to that of the inactivating E211A mutation in Ec Rho (38) which supports that the conserved catalytic glutamate residue (Supplementary Figure S1) is also required in Mtb Rho for polarization of the water molecule attacking the γ -phosphoryl group of ATP.

Taken together, these data establish that coupling between RNA binding and ATP hydrolysis is tight within the Mtb Rho ring and most likely relies on similar principles in the Mtb Rho and Ec Rho enzymes. One should recall, however, that the rate of ATP hydrolysis is a measure of fuel consumption but by no means a measure of motor efficiency. The latter depends on multiple factors governing the mechanochemical transduction mechanism and yield (42), some of which may vary between the Mtb Rho and Ec Rho motors.

Mtb Rho unwinds short RNA duplexes in a directional, ATP-dependent fashion

To probe the motor properties of Mtb Rho, we performed duplex unwinding (helicase) experiments with model substrates (26). We prepared RNA:DNA substrates containing a 57 bp (substrate A) or 34 bp (substrate B) duplex region located downstream from a single-stranded RNA (ssRNA) tail encompassing a synthetic *Rut* site (Figure 3, inset) (30). We also prepared a 5'-ssRNA tailed substrate containing a shorter duplex (14 bp, substrate C) by pairing an oligo(2'-O-methylribonucleotide) to the RNA strand (Figure 3, inset). The 2'-O-methyl modifications, which increase RNA duplex stability significantly (43), prevent spontaneous denaturation of substrate C under our experimental conditions (Figure 3, ADP lanes in bottom gels).

We incubated each substrate with a 10-fold molar excess of Rho hexamers in the presence of ATP under 'standard' helicase reaction conditions (see 'Material and Methods' section) (30,32). As expected, Ec Rho unwinds the three duplex substrates with high efficiencies (Figure 3, left gels and graph). Substrate C is slightly less reactive than the longer duplex substrates, probably because its 2'-O-methyl moieties favor an A-form duplex conformation and non-productive interactions with Ec Rho (31). By contrast,

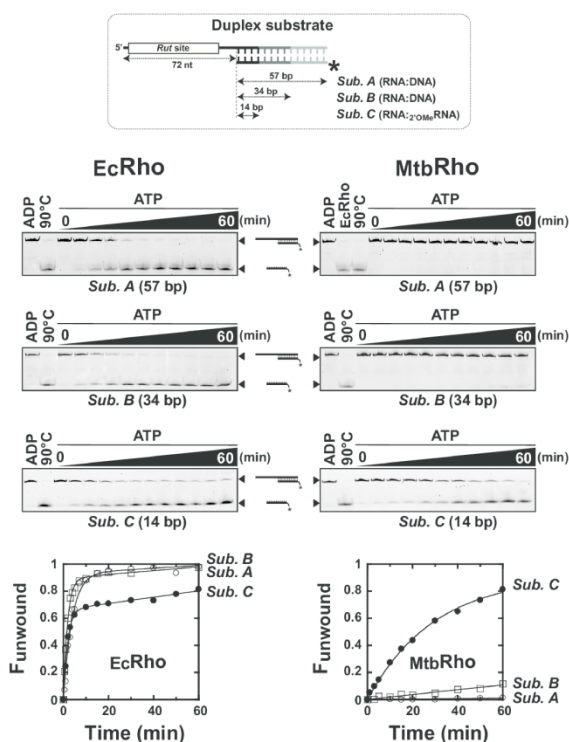


Figure 3. The Mtb Rho helicase unwinds short duplex substrates containing a 5'-single-stranded RNA tail. The duplex substrates used in helicase assays contain a synthetic *Rut* site for Ec Rho (30) and are schematically depicted at the top of the figure. Unwinding reactions were performed under standard conditions with 50 nM of the Ec Rho (left) or Mtb Rho (right) helicase and 5 nM of the indicated substrate. Representative 9% PAGE gels and fraction unwound versus time graphs illustrate the helicase activities observed in each case ($N_R = 3$ except for experiments with Mtb Rho and substrate C [$N_R = 5$]).

Mtb Rho unwinds substrate C most efficiently, unwinds substrate B very inefficiently, and is completely unable to unwind the 57 bp-long duplex of substrate A (Figure 3, right gels and graph). Incubation of the substrates with larger excesses (100-fold) of Mtb Rho does not change unwinding kinetics significantly (data not shown), indicating that Mtb Rho was already present at a saturating concentration. These data demonstrate that Mtb Rho is a RNA helicase, albeit one that works on a narrower range of substrates than Ec Rho.

This narrow spectrum of activity may be due to suboptimal interactions between Mtb Rho and the substrates. Optimal *Rut* features for Mtb Rho are currently unknown and may not be adequately represented by our synthetic *Rut* sequence (Figure 3, inset), which had been optimized for Ec Rho (30,44). Furthermore, the interaction network promoting RNA translocation may be weakened in Mtb Rho by the K $\rightarrow$ T substitution at position 501 of the R-loop (corresponding to SBS residue Lys₃₂₆ in Ec Rho; see above). To test these scenarios, we performed unwinding experiments with Mtb Rho protein variants. We observed that the $[\text{Mtb:Ec}]$ Rho chimera unwinds the three substrates with effi-

Table 1. RNA binding parameters^a

Rho factor	No nucleotides		+ AMP-PNP		+ ATP	
	<i>n</i>	<i>K_d</i> (nM)	<i>n</i>	<i>K_d</i> (nM)	<i>n</i>	<i>K_d</i> (nM)
WT _{Mtb} Rho	1.2 ± 0.2	3.5 ± 0.7	0.9 ± 0.2	2.0 ± 0.3	1.0 ± 0.2	2.4 ± 0.6
E386A	1.5 ± 0.3	5.3 ± 0.8	0.9 ± 0.2	2.2 ± 0.4	1.1 ± 0.4	3.3 ± 1.0
Δ _{Mtb} Rho	2.3 ± 0.6	20.4 ± 2.3	1.4 ± 0.2	16.3 ± 0.4	1.7 ± 0.3	14.4 ± 1.0

^aValues were determined from the best non-linear least square fits of binding data to the Hill equation, $F = F_{\infty} \times [\text{Rho}]^n / (K_d^n + [\text{Rho}]^n)$, where *F* is the fraction of labeled substrate retained on the nitrocellulose membrane, *F_∞* is the maximal fraction of bound substrate, *n* is the Hill coefficient and *K_d* is the apparent dissociation constant. Experiments were performed with the *Rut*-containing RNA strand of helicase substrates (Figure 3) and the indicated _{Mtb}Rho variants as described in methods.

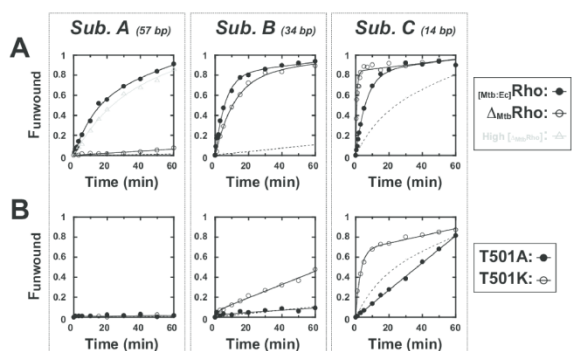


Figure 4. Helicase activities of the _{Mtb}Rho protein variants. Unwinding experiments were performed under standard conditions with substrate A, B or C (Figure 3, inset) and (A) the _{Mtb:Ec}Rho chimera and Δ_{Mtb}Rho deletion mutant or (B) single-point mutants T501A and T501K. To facilitate comparisons, reaction time-courses for WT _{Mtb}Rho are depicted as dotted curves.

ciencies that approach the ones displayed by _{Ec}Rho (Figure 4A, filled circles; compare with Figure 3). The ATPase rates of _{Mtb:Ec}Rho and _{Ec}Rho are also comparable (Figure 2C). These data indicate that _{Mtb}Rho's NTD is reasonably functional (and able to use our synthetic *Rut* sequence) when compared to _{Ec}Rho's NTD. Notwithstanding, deletion of the NTD insertion subdomain (Δ_{Mtb}Rho variant) also significantly increase the capacity of _{Mtb}Rho to unwind the 14 bp and 34 bp duplex substrates (Figure 4A, open circles). When present at a 10-fold higher concentration (500 nM), the Δ_{Mtb}Rho variant can even unwind the 57 bp duplex of substrate A with reasonable efficiency (Figure 4A, open triangles) which, as stated above, is not the case of WT _{Mtb}Rho. These data combined with equilibrium binding measurements (Table 1) confirm that the NTD insertion subdomain stimulates RNA binding but interferes with motor function.

The T501A and T501K mutations have distinct effects on the helicase activity of _{Mtb}Rho. The T501A mutant is slightly less active than the WT enzyme with the shortest (14 bp) duplex substrate and similarly inefficient with the longer duplexes (Figure 4B, filled circles). By contrast, the T501K mutant is significantly more active and able to unwind a longer (34 bp) duplex than WT _{Mtb}Rho (Figure 4B, open circles). Moreover, the amplitude of the exponential phase of the unwinding reaction (with Substrate C) is null for T501A but is ~2-fold higher for T501K than for WT

(Figure 4B). It has been shown for _{Ec}Rho that the amplitude of the exponential phase depends on helicase processivity under similar unwinding reaction conditions (30,32). These observations suggest that stabilizing, productive contacts are formed between the R-loop and RNA within the SBS of _{Mtb}Rho but that these contacts are suboptimal unless Thr₅₀₁ is replaced by a lysine. Significantly, a lysine or arginine is found at the corresponding R-loop position in the Rho factors from most species (e.g. Lys₃₂₆ in _{Ec}Rho) (13), with the notable exception of mycobacteria and coryneform relatives (data not shown). Taken together, these findings indicate that features weakening the _{Mtb}Rho helicase (with respect to the _{Ec}Rho prototype) are located both in its N-terminal 'RNA anchoring' and in its C-terminal 'motor' domains.

Selective unwinding of short duplexes by _{Mtb}Rho is reminiscent of the activity of DEAD-box proteins which disrupt short RNA duplexes in a local, non-directional fashion. However, _{Mtb}Rho is unable to unwind a substrate C derivative wherein the relative positions of the *Rut*-containing ss-RNA and 14bp duplex regions are permuted (Supplementary Figure S3). The _{Ec}Rho enzyme also has no [3'→5'] unwinding activity (30) whereas DEAD-box protein Ded1 unwinds substrate C and its permuted derivative with similar high efficiencies (Supplementary Figure S3), as expected (45). These data thus support that _{Mtb}Rho is a directional [5'→3'] helicase behaving more like the _{Ec}Rho translocase than the local 'disruptase' Ded1.

ATP hydrolysis is required for duplex unwinding by _{Mtb}Rho

A hallmark of RNA translocases is their tight dependence on NTP hydrolysis. By contrast, local disruptases such as DEAD-box proteins often do not require ATP hydrolysis to unwind short RNA duplexes provided that their RNA clamp conformation is stabilized by ATP or a non-hydrolysable ATP analog (46,47). A similar capacity has been proposed for _{Mtb}Rho which is able to disrupt an artificially stalled TEC in the presence of AMP-PNP or ATPγS, albeit at rates that are only, respectively, ~7 and ~23% of the dissociation rate measured in the presence of ATP (25). We performed RNA affinity measurements to test this proposal but observed that _{Mtb}Rho:RNA complexes are only marginally stabilized by nucleotides such as AMP-PNP or ATP (Table 1).

To explore the relationship between the ATPase activity and motor function of _{Mtb}Rho, we also carried out unwinding experiments with the 14 bp duplex substrate C (Figure 3) in the presence of various analogs of ATP. We found that non-hydrolysable ATP analogs AMP-

PNP, ADP-BeF_x (ground-state mimic), and ADP-AlF₄ (transition-state mimic) do not promote unwinding of substrate C by MtbRho (Figure 5A). Similarly, we observed that the ATP γ S analog, which is hydrolyzed by MtbRho at a $\sim$ 15-fold lower rate than ATP under our standard conditions (data not shown), does not stimulate the MtbRho helicase significantly (Figure 5A). UTP and CTP are also highly inefficient helicase cofactors whereas GTP is only slightly less efficient than ATP at promoting unwinding of substrate C by MtbRho (Figure 5A). Overall, the order of NTP efficiency is very similar for duplex unwinding (Figure 5A) and for NTP hydrolysis by MtbRho (Figure 2B).

The catalytic mutant E386A does not unwind substrate C significantly (Figure 5B) which confirms that ATP hydrolysis is required for the helicase function of MtbRho. Further support for this conclusion comes from the observation that BCM inhibits the MtbRho helicase in a dose-dependent fashion (Figure 5C), albeit with a half-maximal inhibitory concentration (IC₅₀ = [157 $\pm$ 20] μ M) that is higher than for EcRho (IC₅₀ = [49 $\pm$ 5] μ M). Altogether, these data support that MtbRho is a *bona fide* ATP-dependent molecular motor which requires ATP hydrolysis to trigger duplex unwinding.

Transcription termination requires a catalytically-competent ATPase pocket in MtbRho

To assess how the peculiar features of MtbRho affect its transcription termination activity, we performed *in vitro* transcription termination experiments with the RNAP of *E. coli* and a DNA template encoding the Rho-dependent terminator λ trI (Figure 6A) (27,34). With this heterologous model system and under single-round ‘chase’ conditions (see ‘Material and Methods’), MtbRho triggers transcription termination with a global apparent efficiency (TE_{app}) that is comparable to that of EcRho (Figure 6B). However, the termination window (i.e. the region encompassing all RNA release sites) is large and starts much closer to the promoter than for EcRho (Figure 6B, compare gel lanes 2 and 3; see also Supplementary Figure S4A) which is consistent with previous observations (25).

Promoter-proximal (early) termination stop points are reminiscent of the behavior of the MlutRho factor from *M. luteus* which also contains a large hydrophilic subdomain in its NTD (Supplementary Figure S1) (40). Deletion of this subdomain in MlutRho impairs early termination (48). We thus tested the termination activity of the Δ MtbRho variant, which lacks the NTD insertion subdomain of MtbRho, and observed that Δ MtbRho has a reduced capacity to trigger early termination (Figure 6B, lane 4; Supplementary Figures S4A and S5). By contrast, the [Mtb:Ec]Rho chimera triggers strong early termination starting even earlier than WT MtbRho (Figure 6B, lane 5; Supplementary Figures S4 and S5).

Importantly, the catalytic mutant E386A is unable to trigger efficient transcript release (Figure 6B, lane 6; Supplementary Figure S4A) supporting that ATP hydrolysis is mandatory for transcription termination induced by MtbRho. By contrast, the T501A mutation only marginally affects transcription termination under our standard conditions by slightly favoring late over early termination (Figure 6B, lane 7; Supplementary Figures S4A and S5). Finally,

the T501K mutant has an increased capacity to trigger early termination but its TE_{app} is not much different from that of WT MtbRho under our standard experimental conditions (Figure 6B, lane 8 and histogram; see also Supplementary Figures S4A and S5).

To confirm the importance of ATP hydrolysis for the termination activity of MtbRho, we performed transcription termination experiments in the presence of increasing concentrations of BCM. We observed that BCM inhibits transcription termination induced by MtbRho in a dose-dependent fashion (Figure 6C; see also Supplementary Figure S4B) but with a global IC₅₀ = [289 $\pm$ 51] μ M that is one order of magnitude higher than for EcRho (IC₅₀ = [27 $\pm$ 3] μ M). Quantification of termination efficiencies at individual termination positions yielded similar results with individual IC₅₀ values ranging between $\sim$ 150 and $\sim$ 300 μ M for MtbRho and being lower than $\sim$ 40 μ M for EcRho (data not shown). Inhibition of the termination activity of the Δ MtbRho mutant also requires high BCM concentrations (global IC₅₀ = [281 $\pm$ 29] μ M; Figure 6D), ruling out that BCM activity is perturbed significantly by the NTD insertion subdomain in MtbRho. By contrast, the [Mtb:Ec]Rho chimera is inhibited at significantly lower BCM concentrations with a global IC₅₀ of [95 $\pm$ 13] μ M (Figure 6D) and IC₅₀ values for individual termination positions as low as $\sim$ 40 μ M (Supplementary Figure S6). Taken together, these data indicate that MtbRho has a low apparent affinity for BCM that is due primarily to the composition of its CTD but confirm nonetheless that MtbRho requires the energy derived from ATP hydrolysis to trigger transcription termination.

Interestingly, the proportion of late (longer than $\sim$ 220 nt) termination products formed with the [Mtb:Ec]Rho chimera tend to increase at low BCM concentrations which contrasts with the evolution of early termination products (Figure 6D and Supplementary Figure S6). This suggests that sub-inhibitory concentrations of BCM trigger a downstream shift of the termination window, possibly because BCM slows down processive translocation of the [Mtb:Ec]Rho chimera and alters its kinetic competition with RNAP (49). This phenomenon is not detected for WT MtbRho (Figure 6C and data not shown), which seems consistent with its low apparent processivity deduced from helicase experiments (compare Figures 3 and 4A). Hence, BCM-impaired MtbRho may not remain bound to the nascent transcript for sufficiently long times (and translocated distances) to catch up with RNAP at alternative pausing sites located further away from the initial *Rut* loading site.

DISCUSSION

The mtbRho factor is a *bona fide* molecular motor

Transcription termination factor Rho is widespread in bacteria (13) where it supports important physiological functions linked to Rho-induced disruption of TECs or R-loops (7–9). Many Gram-negative species are sensitive to the antibiotic BCM (9), a naturally occurring, non-competitive inhibitor of Rho’s ATPase (20), whereas only one tested Gram-positive species—*M. luteus*—has been found to be sensitive to BCM (16). This spectrum of BCM activity suggested that Rho is dispensable in most Gram-positive bac-

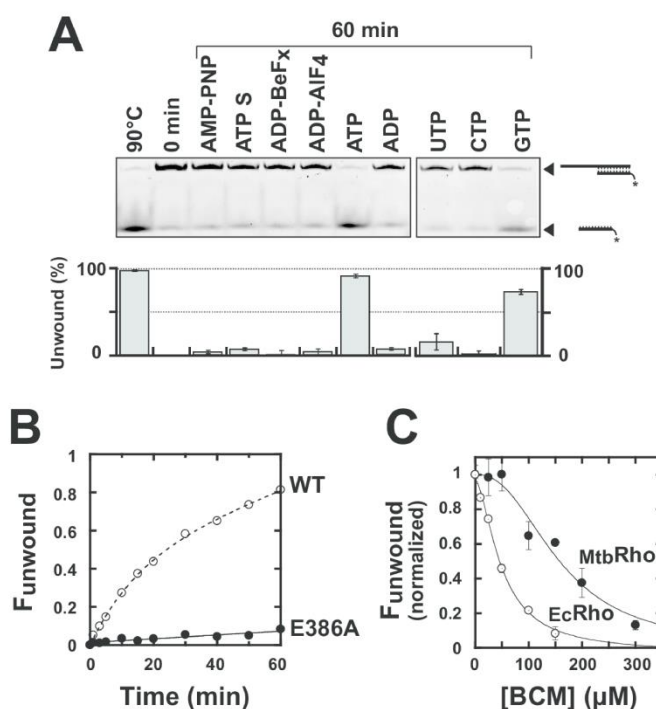


Figure 5. ATP hydrolysis is required for duplex unwinding by *MtbRho*. Experiments were performed with substrate C under standard helicase conditions unless specified otherwise. (A) Comparison of the activation potential of various nucleotide cofactors (each present at a final concentration of 1 mM). (B) The ATPase-deficient mutant E386A does not unwind substrate C efficiently. (C) BCM-induced inhibition of the *EcRho* (open circles) and *MtbRho* (filled circles) helicases. The graph shows the normalized fractions of unwound duplex C formed after 60 min of incubation in the presence of Rho and ATP as a function of the concentration of BCM present in the reaction mixture. Note that BCM has complex effects on the kinetic regimens of the multirun helicase reactions (not shown) which prevent determination of reliable inhibitory parameters based on the evolution of unwinding reaction amplitudes or rates as a function of BCM concentration.

teria, a proposal that has been tested and verified for only a few specimens such as *S. aureus* (17). Recently, an alternative explanation has been proposed for mycobacterial resistance to BCM (25). Based on the observation that *MtbRho* disrupted stalled TECs in the presence of AMP-PNP or ATP γ S and because the authors failed to detect a significant helicase activity for *MtbRho*, they proposed that the enzyme is able to trigger transcription termination by a mechanism uncoupled from molecular motor function (25). Such a mechanism would make *MtbRho* intrinsically resistant to BCM, which is what the authors observed for BCM concentrations up to 200 μ M (25). Our data, however, do not corroborate this hypothesis and suggest alternative explanations as is discussed below.

Tight RNA binding by a protein is sometimes sufficient to destabilize proximal complexes on RNA. This principle is utilized by DEAD-box proteins which bind ATP and RNA cooperatively to disrupt short RNA duplexes (or RNA:protein complexes) in a non-directional fashion (1,6). However, *MtbRho* does not bind RNA much tighter in the presence of nucleotides (Table 1) and differs from DEAD-box proteins by two other important criteria. First, *MtbRho* is unable to unwind short duplex substrates in the presence of non-hydrolysable ATP analogs (Figure 5A). Second,

MtbRho cannot unwind substrates wherein the duplex region is located upstream of the ssRNA loading (e.g. *Rut*) site (Supplementary Figure S3). Rather, the *MtbRho* enzyme behaves as a *bona fide* ATP-dependent [5'→3'] RNA translocase only able to unwind short RNA duplexes located downstream from a *Rut* site (Figure 3). This activity strictly requires ATP (or GTP) hydrolysis (Figure 5B), a condition confirmed by the inhibition of duplex unwinding induced by the non-conservative mutation of the catalytic Glu₃₈₆ residue (Figure 5B, E386A mutation) or by high concentrations of BCM (Figure 5C). Importantly, *in vitro* transcription termination by *MtbRho* is similarly inhibited by the E386A mutation (Figure 6B) or by high BCM concentrations (Figure 6C), supporting that both the helicase and termination functions of *MtbRho* proceed by the same motor-dependent mechanism. This conclusion is in apparent contradiction with the previous observation that artificially stalled TECs can be dissociated by *MtbRho* in the presence of ATP γ S or AMP-PNP (25). However, the reported TEC dissociation rates were low, especially in the presence of AMP-PNP (~7% of the ATP-dependent rate) (25), and may thus reflect residual motor-independent destabilization as much as experimental limitation(s) due, for instance, to slow hydrolysis of ATP γ S by *MtbRho* (see 'Results' section)

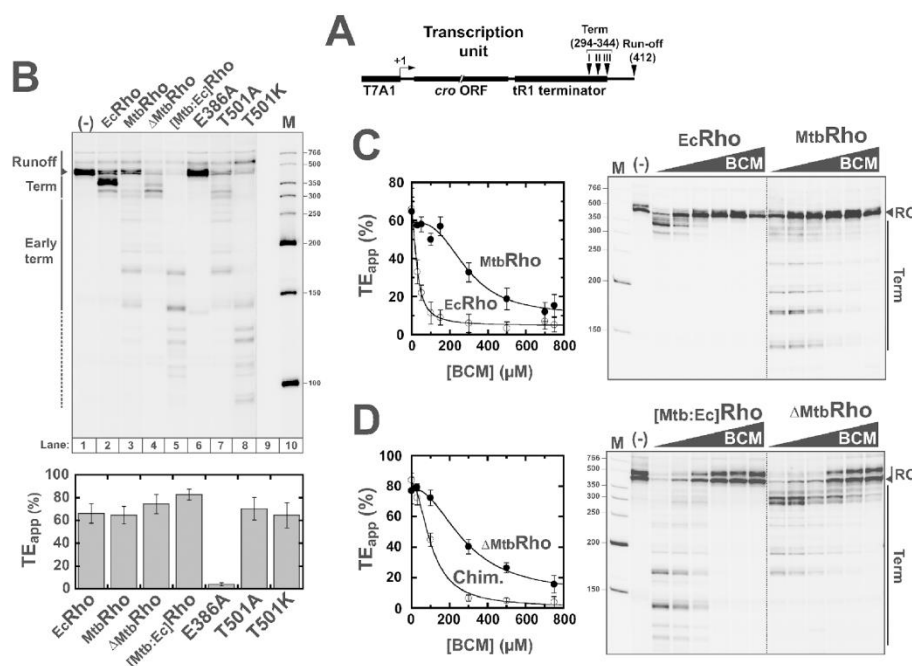


Figure 6. Transcription termination experiments. (A) Schematic of the DNA template containing the λ R1 terminator used in *in vitro* termination experiments. (B) Both the NTD and CTD of $MtbRho$ affect its transcription termination activity. A representative denaturing PAGE gel showing the products of transcription termination experiments performed under standard 'chase' conditions with *E. coli* RNAP and the λ R1 terminator in the presence of various Rho factors (indicated above gel lanes). Intensity of the T501K lane had to be adjusted separately (as indicated by vertical dotted lines) because reactions with the T501K mutant consistently yielded lower product amounts. Apparent termination efficiencies (TE_{app}) are indicated below the gel ($N_R = 3$). (C) A higher concentration of BCM is required to inhibit $MtbRho$ than $EcRho$ under standard transcription termination 'chase' conditions ($N_R = 3$). (D) Comparison of the termination activities of the $[Mtb:Ec]Rho$ and $\Delta MtbRho$ variants in the presence of BCM indicates that structural components conferring BCM resistance to $MtbRho$ are located mostly in its CTD. In the figure, the intensity and gamma settings were adjusted to facilitate the comparison of termination signals in the various gel panels (separated by dotted lines). Gel images without gamma setting adjustments are presented in Supplementary Figures S6 and S7E.

or contamination of commercial AMP-PNP preparations by ATP (50).

Limited helicase processivity versus high termination efficiency of *mtbRho*

Although our data unambiguously demonstrate that $MtbRho$ is a directional, ATP-fueled RNA helicase, they also show that the enzyme is much less efficient than its $EcRho$ homolog at unwinding long duplexes (Figure 3). This low processivity of $MtbRho$ derives from at least two specific structural features which are discussed below.

First, the presence of a threonine residue (instead of a basic lysine/arginine residue, as in most Rho homologs) (13) at the C-terminal position of the R-loop (Thr₅₀₁) most likely weakens the SBS:RNA interaction network in $MtbRho$. Replacement of Thr₅₀₁ by a lysine (T501K mutant) enhances the capacity of $MtbRho$ to unwind longer duplexes whereas its replacement by alanine (T501A mutant) abrogates the unwinding burst observed with a short duplex (Figure 4B). These observations are consistent with Rho helicase processivity depending on optimal SBS:RNA interactions (11). Interestingly, the T501K mutation also strongly favors early termination whereas T501A slightly disfavors it (Figure 6B;

Supplementary Figure S5). The initial capture of a *Rut* site, which depends on PBS:RNA contacts (7–9), should not be affected by modifications of SBS residues. However, subsequent activation of the *Rut*-bound Rho motor may be stimulated by extra (or stronger) SBS:RNA contacts, thereby allowing earlier catch up with RNAP. The superior behavior of the $[Mtb:Ec]Rho$ chimera (Figures 4A and 6B), which contains the SBS of $EcRho$, lends support to this proposal. One should also note that the suboptimal R-loop configuration in $MtbRho$ probably increases the dependency of the SBS:RNA network on Q-loop contacts. This conjecture is supported by preliminary analysis of Q-loop mutant phenotypes (our unpublished observations). For instance, the single-point S461A mutant of $MtbRho$ displays a strong (~10-fold) defect in poly[rC]-dependent ATPase activity (Supplementary Figure S7) whereas the corresponding T286A mutation in the Q-loop of $EcRho$ only reduces ATPase activity by 35% (51). This difference should be considered in light of the fact that Thr₂₈₆ side chains make direct contacts with RNA in the crystal structure of $EcRho$ (10).

Second, the presence of a large, hydrophilic NTD insertion subdomain in $MtbRho$ (Figure 1B) appears to perturb

the coupling between RNA binding and motor functions. Deletion of this subdomain (Δ_{MtbRho} variant) weakens RNA binding (Table 1) but significantly increases the ATPase turnover and apparent duplex unwinding processivity of MtbRho at saturating RNA concentrations (see 'Results' section and Figure 4A). Bulky NTD subdomain appendages (Figure 1C) may hamper the structural rearrangements driving enzyme activation and mechanochemical transduction (i.e. transformation of chemical energy into mechanical work) in the Rho ring (10,52). Such constraints are, however, much less detrimental in the $[\text{Mtb:Ec}]$ Rho chimera which displays processive unwinding (Figure 4A) and strong termination (Figure 6B) activities. The NTD subdomain thus appears to specifically exacerbate the sub-optimal behavior of the MtbRho motor.

We cannot totally exclude that the processivity of the MtbRho helicase is also mitigated by suboptimal PBS interactions with the duplex substrates. Our synthetic *Rut* sequence has been optimized for EcRho (30,44) and may thus not adequately recapitulate the distinct features of *M. tuberculosis* transcripts (39). An optimal $\text{MtbRho}:\text{Rut}$ interaction crowning the Rho hexamer could enhance intersubunit cohesion (and ring stability) during tethered tracking, thereby favoring further RNA translocation over RNA exit from the Rho ring and dissociation (11,12). Comparison of the enzymatic activities of the MtbRho and $[\text{Mtb:Ec}]$ Rho factors (Figures 2C, 4A and 6) nonetheless supports that EcRho contains a stronger, more processive motor than MtbRho . It thus remains to be seen whether MtbRho will be able to disrupt stable R-loop structures under appropriate conditions. This capacity could have some value if MtbRho has to deal with transcriptional R-loops *in vivo*, as is the case for EcRho (Figure 1A) (53), which is not known yet.

Although MtbRho efficiently terminates transcription by *E. coli*'s RNAP under conditions that favor fast RNAP translocation (Figure 6A; Supplementary Figure S4), our data suggest that MtbRho cannot operate far from its loading site on the nascent transcript. This proposal is notably supported by the lack of downstream shift of the MtbRho termination window in the presence of sub-inhibitory concentrations of BCM (see 'Results' section). We thus propose that, while the processive EcRho has evolved to catch up with RNAP at sites located up to ~130 nt downstream from the *Rut* region (7–9), MtbRho needs to promptly inactivate its RNAP target. This task could be facilitated by the significantly lower elongation rate of *M. tuberculosis*' RNAP (54) or by direct interactions between Rho and RNAP (55) that could be stronger in the case of MtbRho (e.g. due to the NTD insertion subdomain). Further work is necessary to rigorously test these intriguing possibilities.

Origin(s) of BCM resistance

The low apparent affinity of MtbRho for BCM that was deduced from helicase (Figure 5C) and transcription termination (Figure 6B and C) experiments is intriguing given that MtbRho does not contain mutations that are known to confer BCM resistance (13). These mutations usually disrupt molecular contacts between amino acids and BCM in the ATPase pocket of Rho, thereby destabilizing the Rho:BCM complex (20). We could not determine BCM affinities by

microcalorimetry because of experimental constraints associated with weak ligand binding (56) and foaming of concentrated MtbRho solutions in the microcalorimeter mixing chamber (our unpublished observations). Thus, weakening of BCM binding by unknown structural determinants in MtbRho (e.g. ones that would alter the shape of, or access to, the ATPase pocket) cannot be formally ruled out. Other explanations based on distinctive mechanochemical coupling in MtbRho are also possible, although their rigorous testing is beyond the scope of the present study.

Conserved transcription termination mechanisms in *M. tuberculosis*

Pathogenic mycobacteria have long adapted their metabolisms and genetic programs to the colonization of harsh and changing environments such as host macrophages and granuloma (57). How this adaptation is enforced at the level of mycobacterial transcription is only fragmentarily understood. The two major pathways leading to transcription termination in bacteria—that is, intrinsic termination and Rho-dependent termination—have been proposed to be markedly distinct in *M. tuberculosis* (25,58). However, a recent study demonstrated that mycobacterial RNAP terminates transcription at canonical intrinsic terminators in a process which can be aided by NusG, as in other bacteria (54). Mycobacterial RNAP only displayed an increased capacity (as compared to *E. coli*'s RNAP) to use terminators with suboptimal U-tracts (54). In the present work, we demonstrate that MtbRho operates by a canonical 'molecular motor' mechanism that resembles the one used by EcRho . It thus appears that both intrinsic and factor-dependent termination mechanisms are not fundamentally different in *M. tuberculosis* but adjusted to the characteristics of the bacterium that include a GC-rich genome and slow metabolism. Since MtbRho is expressed in *M. tuberculosis* (23) and appears to be essential for growth (21,22), seeking MtbRho inhibitors by using a screening strategy derived from ATP-dependent motor assays (such as the ones presented in this study) could thus prove useful in the fight against tuberculosis, a major and global health concern.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge R. Sen, J.P. Richardson and J.M. Berger for the gift of materials; M. Rabhi, M. Chierif and G. Tanguy for their assistance during early stages of this project; and members of our groups at CBM for helpful discussions and critical reading of the manuscript. We thank the «Plateforme de Spectrométrie de Masse et Protéomique» of CBM for mass spectrometry analyses.

FUNDING

University of Orléans and Région Centre [F.d.H.]; CNRS [B.C., M.B]; French Agence Nationale de la Recherche

[2010-BLAN-152502 to M.B.]. Funding for open access charge: CNRS; French Agence Nationale de la Recherche [2010-BLAN-152502 to M.B.].
Conflict of interest statement. None declared.

REFERENCES

- Jarmoskaite, I. and Russell, R. (2014) RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.*, **83**, 697–725.
- Jankowsky, E. and Fairman-Williams, M.E. (2010) An introduction to RNA helicases: superfamilies, families, and major themes. In: Jankowsky, E. (ed). *RNA Helicases*. RSC Publishing, Cambridge, Vol. 19, pp. 1–31.
- Steimer, L. and Klostermeier, D. (2012) RNA helicases in infection and disease. *RNA Biol.*, **9**, 751–771.
- Rabhi, M., Tuma, R. and Boudvillain, M. (2010) RNA remodeling by hexameric RNA helicases. *RNA Biol.*, **7**, 655–666.
- Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
- Linder, P. and Jankowsky, E. (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.*, **12**, 505–516.
- Boudvillain, M., Figueroa-Bossi, N. and Bossi, L. (2013) Terminator still moving forward: expanding roles for Rho factor. *Curr. Opin. Microbiol.*, **16**, 118–124.
- Peters, J.M., Vangeloff, A.D. and Landick, R. (2011) Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.*, **412**, 793–813.
- Rabhi, M., Rahmouni, A.R. and Boudvillain, M. (2010) Transcription termination factor Rho: a ring-shaped RNA helicase from bacteria. In: Jankowsky, E. (ed). *RNA Helicases*. RSC Publishing, Cambridge (UK), Vol. 19, pp. 243–271.
- Thomsen, N.D. and Berger, J.M. (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell*, **139**, 523–534.
- Soares, E., Schwartz, A., Nollmann, M., Margeat, E. and Boudvillain, M. (2014) The RNA-mediated, asymmetric ring regulatory mechanism of the transcription termination Rho helicase decrypted by time-resolved Nucleotide Analog Interference Probing (trNAIP). *Nucleic Acids Res.*, **42**, 9270–9284.
- Gocheva, V., Le Gall, A., Boudvillain, M., Margeat, E. and Nollmann, M. (2015) Direct observation of the translocation mechanism of transcription termination factor Rho. *Nucleic Acids Res.*, **43**, 2367–2377.
- D'Heygere, F., Rabhi, M. and Boudvillain, M. (2013) Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology*, **159**, 1423–1436.
- OIE. (2014) List of antimicrobial agents of veterinary importance. *World Organization for Animal Health, Annual Report*, pp. 1–8.
- Kohn, H. and Widger, W. (2005) The molecular basis for the mode of action of bicyclomycin. *Curr. Drug Targets Infect. Disord.*, **5**, 273–295.
- Nowatzke, W.L., Keller, E., Koch, G. and Richardson, J.P. (1997) Transcription termination factor Rho is essential for *Micrococcus luteus*. *J. Bacteriol.*, **179**, 5238–5240.
- Washburn, R.S., Marra, A., Bryant, A.P., Rosenberg, M. and Gentry, D.R. (2001) rho is not essential for viability or virulence in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, **45**, 1099–1103.
- Van Bambeke, F., Glupczynski, Y., Plesiat, P., Pechere, J.C. and Tulkens, P.M. (2003) Antibiotic efflux pumps in prokaryotic cells: occurrence, impact on resistance and strategies for the future of antimicrobial therapy. *J. Antimicrob. Chemother.*, **51**, 1055–1065.
- Magyar, A., Zhang, X., Kohn, H. and Widge, W.R. (1996) The antibiotic bicyclomycin affects the secondary RNA binding site of *Escherichia coli* transcription termination factor Rho. *J. Biol. Chem.*, **271**, 25369–25374.
- Skordalakes, E., Brogan, A.P., Park, B.S., Kohn, H. and Berger, J.M. (2005) Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure*, **13**, 99–109.
- Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J. and Sassetti, C.M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathogens*, **7**, e1002251.
- Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.*, **48**, 77–84.
- Wolfe, L.M., Veeraraghavan, U., Idicula-Thomas, S., Schurer, S., Wennerberg, K., Reynolds, R., Besra, G.S. and Dobos, K.M. (2013) A chemical proteomics approach to profiling the ATP-binding proteome of *Mycobacterium tuberculosis*. *Mol. Cell. Proteomics*, **12**, 1644–1660.
- Nishida, M., Mine, Y., Matsubara, T., Goto, S. and Kuwahara, S. (1972) Bicyclomycin, a new antibiotic. 3. In vitro and in vivo antimicrobial activity. *J. Antibiot. (Tokyo)*, **25**, 582–593.
- Kalarickal, N.C., Ranjan, A., Kalyani, B.S., Wal, M. and Sen, R. (2010) A bacterial transcription terminator with inefficient molecular motor action but with a robust transcription termination function. *J. Mol. Biol.*, **395**, 966–982.
- Boudvillain, M., Walmacq, C., Schwartz, A. and Jacquinot, F. (2010) Simple enzymatic assays for the in vitro motor activity of transcription termination factor Rho from *Escherichia coli*. *Methods Mol. Biol.*, **587**, 137–154.
- Rabhi, M., Gocheva, V., Jacquinot, F., Lee, A., Margeat, E. and Boudvillain, M. (2011) Mutagenesis-based evidence for an asymmetric configuration of the ring-shaped transcription termination factor Rho. *J. Mol. Biol.*, **405**, 497–518.
- D'Heygere, F., Schwartz, A., Coste, F., Castaing, B. and Boudvillain, M. (2015) Monitoring RNA Unwinding by the Transcription Termination Factor Rho from *Mycobacterium tuberculosis*. *Methods Mol. Biol.*, **1259**, 293–311.
- Gan, E. and Richardson, J.P. (1999) ATP and other nucleotides stabilize the Rho-mRNA complex. *Biochemistry*, **38**, 16882–16888.
- Walmacq, C., Rahmouni, A.R. and Boudvillain, M. (2004) Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. *J. Mol. Biol.*, **342**, 403–420.
- Schwartz, A., Walmacq, C., Rahmouni, A.R. and Boudvillain, M. (2007) Noncanonical interactions in the management of RNA structural blocks by the transcription termination rho helicase. *Biochemistry*, **46**, 9366–9379.
- Walmacq, C., Rahmouni, A.R. and Boudvillain, M. (2006) Testing the steric exclusion model for hexameric helicases: substrate features that alter RNA-DNA unwinding by the transcription termination factor Rho. *Biochemistry*, **45**, 5885–5895.
- Figueroa-Bossi, N., Schwartz, A., Guillemardet, B., D'Heygere, F., Bossi, L. and Boudvillain, M. (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev.*, **28**, 1239–1251.
- Rabhi, M., Espeli, O., Schwartz, A., Cayrol, B., Rahmouni, A.R., Arluison, V. and Boudvillain, M. (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *Embo J.*, **30**, 2805–2816.
- Burton, M.E. (2006) *Applied pharmacokinetics & pharmacodynamics: principles of therapeutic drug monitoring*, 4th edn. Lippincott Williams & Wilkins, Baltimore, MD.
- Canals, A., Usón, I. and Coll, M. (2010) The structure of RNA-free Rho termination factor indicates a dynamic mechanism of transcript capture. *J. Mol. Biol.*, **400**, 16–23.
- Skordalakes, E. and Berger, J.M. (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell*, **114**, 135–146.
- Balasubramanian, K. and Stitt, B.L. (2010) Evidence for amino acid roles in the chemistry of ATP hydrolysis in *Escherichia coli* Rho. *J. Mol. Biol.*, **404**, 587–599.
- Mitra, A., Misquitta, R. and Nagaraja, V. (2014) *Mycobacterium tuberculosis* Rho is an NTPase with distinct kinetic properties and a novel RNA-binding subdomain. *PLoS One*, **9**, e107474.
- Nowatzke, W.L. and Richardson, J.P. (1996) Characterization of an unusual Rho factor from the high G + C Gram-positive bacterium *Micrococcus luteus*. *J. Biol. Chem.*, **271**, 742–747.
- Ingham, C.J., Hunter, I.S. and Smith, M.C. (1996) Isolation and sequencing of the rho gene from *Streptomyces lividans* ZX7 and characterization of the RNA-dependent NTPase activity of the overexpressed protein. *J. Biol. Chem.*, **271**, 21803–21807.

42. Bustamante, C., Cheng, W. and Mejia, Y.X. (2011) Revisiting the central dogma one molecule at a time. *Cell*, **144**, 480–497.
43. Cummins, L.L., Owens, S.R., Risen, L.M., Lesnik, E.A., Freier, S.M., McGee, D., Guinasso, C.J. and Cook, P.D. (1995) Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. *Nucleic Acids Res.*, **23**, 2019–2024.
44. Guerin, M., Robichon, N., Geiselmann, J. and Rahmouni, A.R. (1998) A simple polypyrimidine repeat acts as an artificial Rho-dependent terminator in vivo and in vitro. *Nucleic Acids Res.*, **26**, 4895–4900.
45. Yang, Q. and Jankowsky, E. (2006) The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Mol. Biol.*, **13**, 981–986.
46. Liu, F., Putnam, A. and Jankowsky, E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20209–20214.
47. Liu, F., Putnam, A.A. and Jankowsky, E. (2014) DEAD-box helicases form nucleotide-dependent, long-lived complexes with RNA. *Biochemistry*, **53**, 423–433.
48. Nowatzke, W.L., Burns, C.M. and Richardson, J.P. (1997) Function of the novel subdomain in the RNA binding domain of transcription termination factor Rho from *Micrococcus luteus*. *J. Biol. Chem.*, **272**, 2207–2211.
49. Jin, D.J., Burgess, R.R., Richardson, J.P. and Gross, C.A. (1992) Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 1453–1457.
50. Meyer, A.S., Gillespie, J.R., Walther, D., Millet, I.S., Doniach, S. and Frydman, J. (2003) Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell*, **113**, 369–381.
51. Wei, R.R. and Richardson, J.P. (2001) Mutational changes of conserved residues in the Q-loop region of transcription factor Rho greatly reduce secondary site RNA-binding. *J. Mol. Biol.*, **314**, 1007–1015.
52. Stitt, B.L. and Xiao, H. (2010) Conformation changes in E. coli Rho monitored by hydrogen/deuterium exchange and mass spectrometry: response to ligand binding. *J. Mol. Biol.*, **402**, 813–824.
53. Krishna Leela, J., Syeda, A.H., Anupama, K. and Gowrishankar, J. (2013) Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 258–263.
54. Czyz, A., Mooney, R.A., Iaconi, A. and Landick, R. (2014) Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio*, **5**, e00931.
55. Epshtein, V., Dutta, D., Wade, J. and Nudler, E. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature*, **463**, 245–249.
56. Brogan, A.P., Widger, W.R., Bensadek, D., Riba-Garcia, I., Gaskell, S.J. and Kohn, H. (2005) Development of a technique to determine bicyclomycin-rho binding and stoichiometry by isothermal titration calorimetry and mass spectrometry. *J. Am. Chem. Soc.*, **127**, 2741–2751.
57. Cambier, C.J., Falkow, S. and Ramakrishnan, L. (2014) Host evasion and exploitation schemes of mycobacterium tuberculosis. *Cell*, **159**, 1497–1509.
58. Unniraman, S., Prakash, R. and Nagaraja, V. (2001) Alternate paradigm for intrinsic transcription termination in eubacteria. *J. Biol. Chem.*, **276**, 41850–41855.

SUPPLEMENTARY METHODS

Circular Dichroism. CD spectra were recorded at 20 °C on a Jasco J-810 spectropolarimeter from 260 to 195 nm using a 1-mm quartz cell, a scan rate of 50 nm/min, a response time of 8 ms, a bandwidth of 1 nm, a resolution of 1 nm, and a protein concentration of 0.1 mg/ml in a buffer containing 10 mM sodium phosphate, pH 7.5 and 100 mM NaCl. Each spectrum represents the average of five scans, with the buffer subtracted.

Diffusion light scattering. Hydrodynamic radius measurements were made at 20 °C with a Zetasizer Nano S instrument (Malvern Instruments Ltd, UK) and a protein concentration of 1mg/ml in GF buffer.

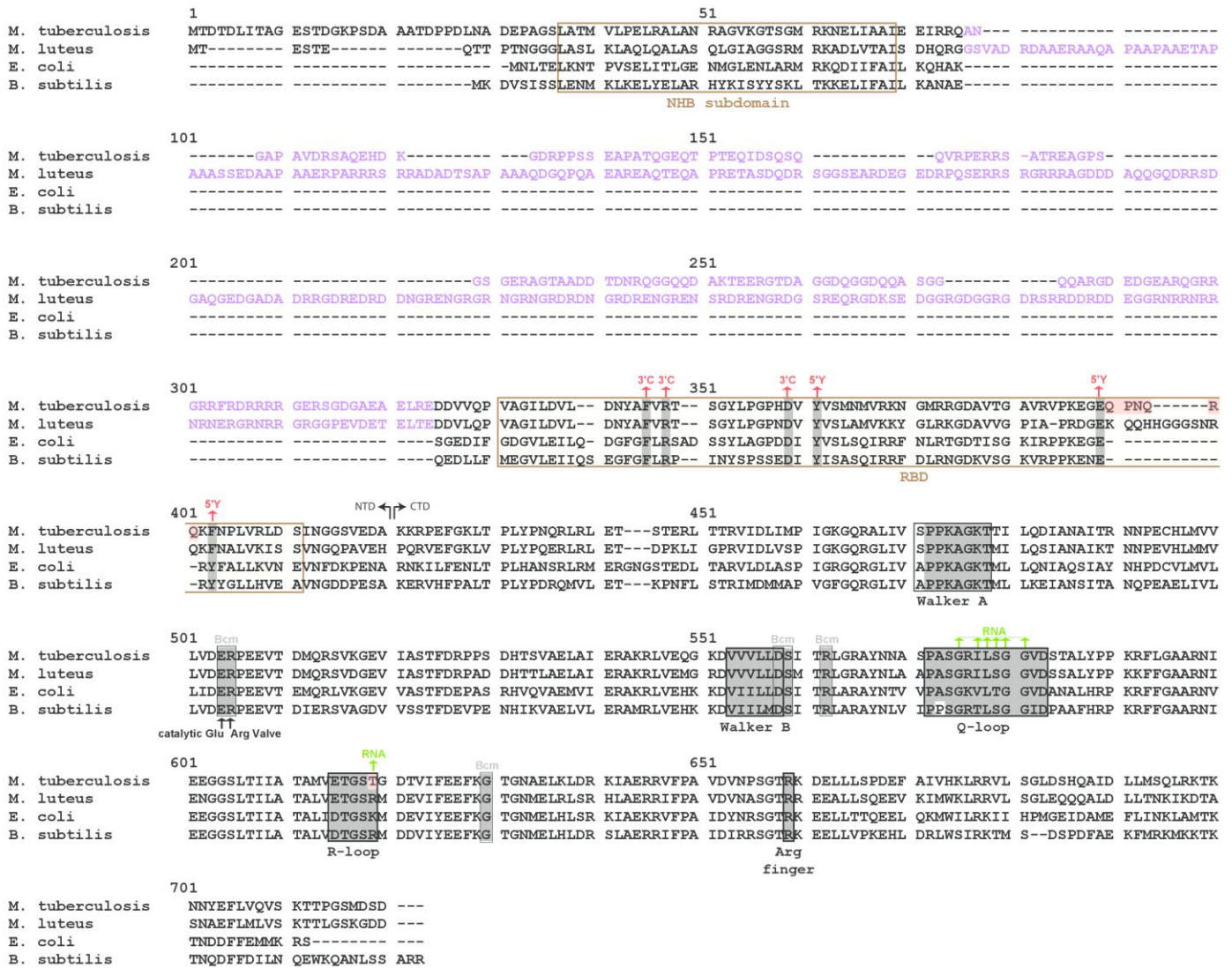
Supplementary Table 1: primers used for plasmid constructs

Plasmid	Method	Source DNA	Forward primer	Reverse primer
pET28b-MtbRho/T501A	Quickchange ^a	pET28b-MtbRho	5'-GTCGAGACCGGGTCCGCTGGTGACACGGTC'	5'-GACCGTGTACACAGCGGACCCGGTCTCGAC
pET28b-MtbRho/T501K	Quickchange ^a	pET28b-MtbRho	5'-GGTCGAGACCGGGTCCAAAGGTGACACGGTCATTTTC	5'-GAAATGACCGTGTACCTTTGACCCGGTCTCGACC
pET28b-MtbRho/E386A	Quickchange ^a	pET28b-MtbRho	5'-GTGCTCGTCGACGCGCGGCTGAGGAG	5'-CTCCTCAGGCCGCGCTCGACGAGCAC
pET28b-MtbRho/S461A	Quickchange ^a	pET28b-MtbRho	5'-GGCCGGATCCTGGCCGGTGGTGTCG	5'-CGACACCACCGGCCAGGATCCGGCC
pET28b-ΔMtbRho	1 [1-75] ^b	pET28b-MtbRho	5'-GGGAATTGTGAGCGGATAAC	5'-CCGGCTGGACGACGTCGTCGGGTACCTGTCGCTGATCTCC
	2 [220-602] ^b	pET28b-MtbRho	5'-GGAGATCAGGCGACAGGTACCCGACGACGTCGTCACGCCGG	5'-GGCCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGC
pET28b-[Mtb:Ec]Rho	1 [1-308] ^b	pET28b-MtbRho	5'-GGGAATTGTGAGCGGATAAC	5'-GGGTAAAGTTCTCAAAGAGGATCCGCTTCTGGCGTCTTCGACCG
	2 [131-419] ^b	pET28b-EcRho	5'-CGGTCGAAGACGCCAAGAAGCGGATCCTCTTGAGAACTTAACCC	5'-GGCCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGC
pET28b-[Ec:Mtb]Rho	1 [1-130] ^b	pET28b-EcRho	5'-GGGAATTGTGAGCGGATAAC	5'-CGTCAGTTTGCCGAACTCGGGTTTGTGCGGGCGTTTTCAGG
	2 [309-602] ^b	pET28b-MtbRho	5'-CCTGAAAACGCCGCAACAAACCCGAGTTCGGCAAACTGACG	5'-GGCCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGC

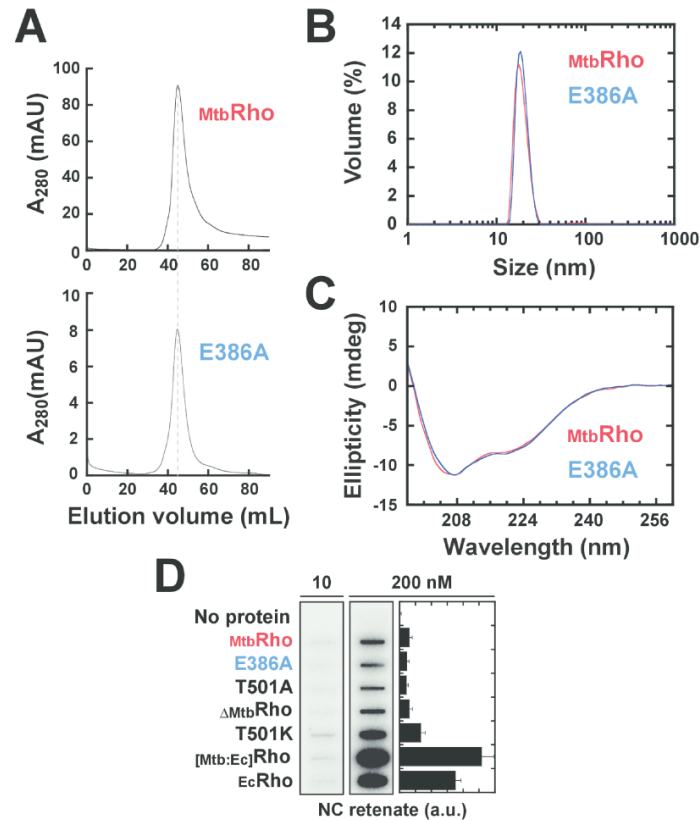
^a Site-directed mutagenesis using the Quickchange method (Stratagene) and the indicated mutagenic primers.

^b Overhang extension PCR and standard cloning. Full insert DNA fragments were obtained by fusing the N-terminal fragment 1 to the C-terminal fragment 2 (corresponding Rho amino acids indicated in brackets) by overhang extension PCR. The insert DNA fragments were cloned into the pET28b vector plasmid after digestion of both vector and inserts with the XbaI and XhoI restriction enzymes.

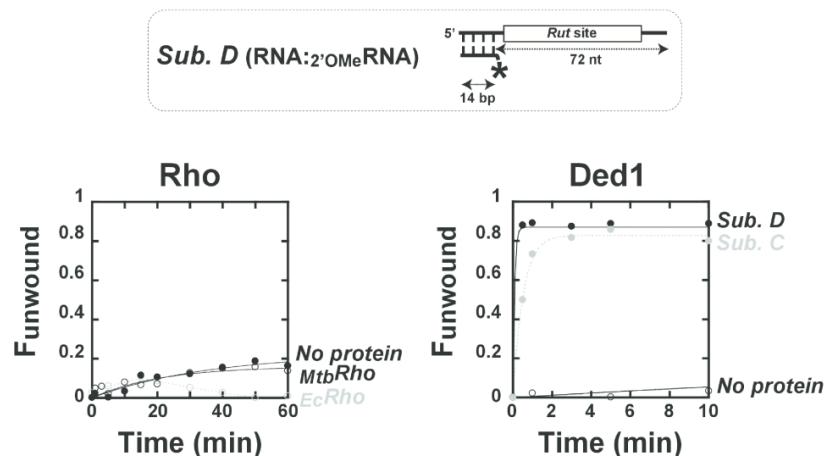
SUPPLEMENTRY FIGURES



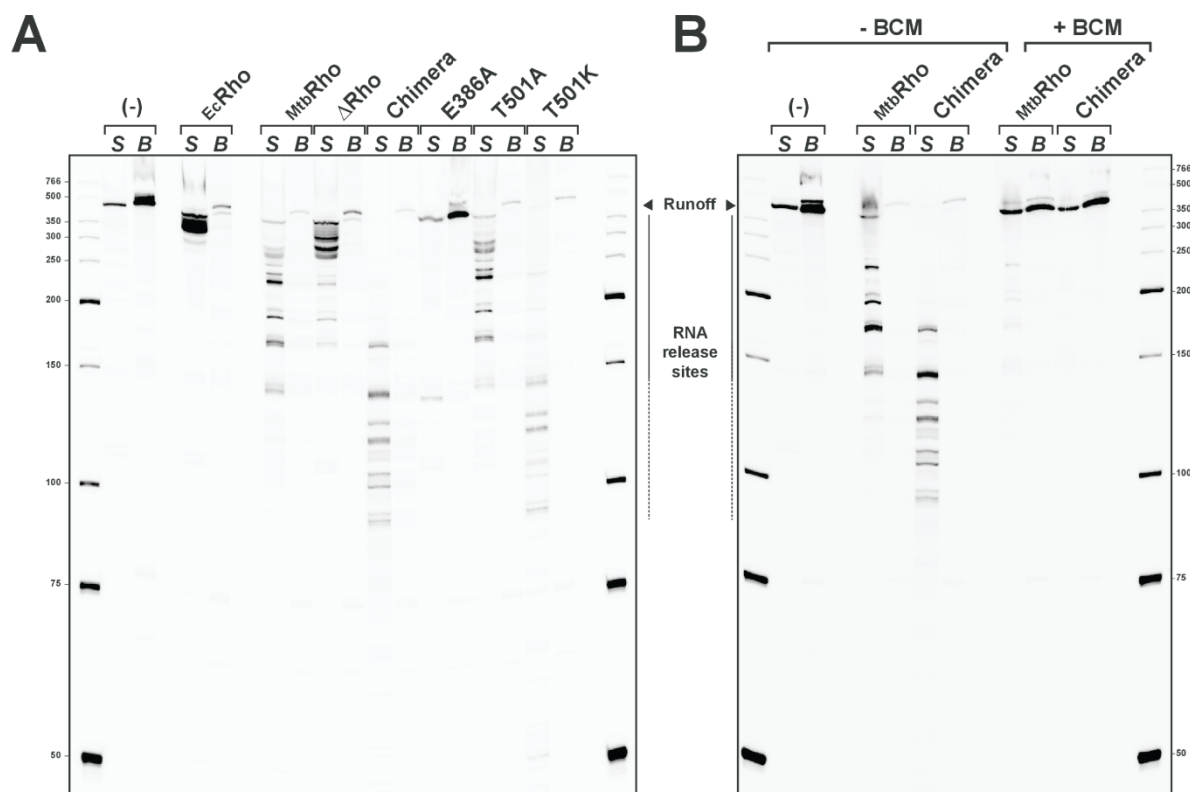
Supplementary figure 1: Alignment of representative Rho sequences (excerpt from Ref. 13). Color coding of major motifs is the same as in Figure 1. Key conserved residues are highlighted in grey. The small PBS insert (aa 281-286) and T501 residue (in R-loop) in *Mtb* Rho are highlighted in light red. Numbers above sequences correspond to positions in the alignment (not in individual sequences).



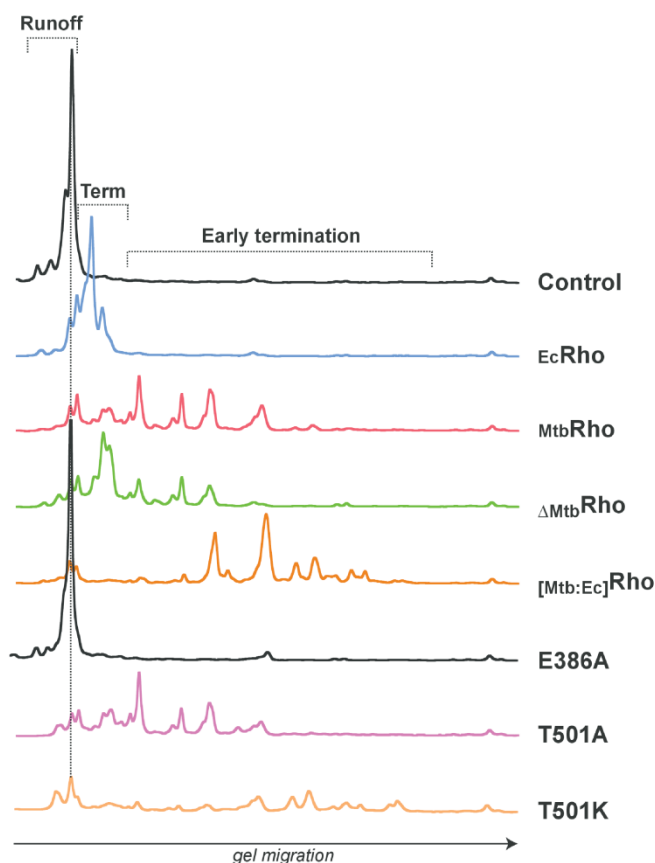
Supplementary figure 2: Single-point mutant E386A and WT MtbRho have comparable structural, oligomerization, and ATP binding behaviors. **(A)** Chromatograms of the E386A and WT MtbRho proteins analyzed by gel filtration on a Sephacryl S-300 HR column. **(B)** Diffusion light scattering and **(C)** circular dichroism spectra of the E386A and WT MtbRho proteins. **(D)** Retention of $[\gamma^{32}\text{P}]\text{ATP}:\text{Rho}$ complexes on a nitrocellulose (NC) membrane. Rho proteins (10 or 200 nM) were incubated with 0.1 nM $[\gamma^{32}\text{P}]\text{ATP}$ in helicase buffer supplemented with 1 mM MgCl_2 for 10 min at 30°C before being filtered through a NC membrane using a slot-blot apparatus. Note that the amounts of ATP retained with proteins on the NC membrane are significantly higher for EcRho and [Mtb:Ec]Rho. This suggests that the EcRho motor binds ATP much tighter than the MtbRho motor and is consistent with the ~10-fold lower $K_{\text{m,ATP}}$ measured for ATP hydrolysis by EcRho than by MtbRho (see main text).



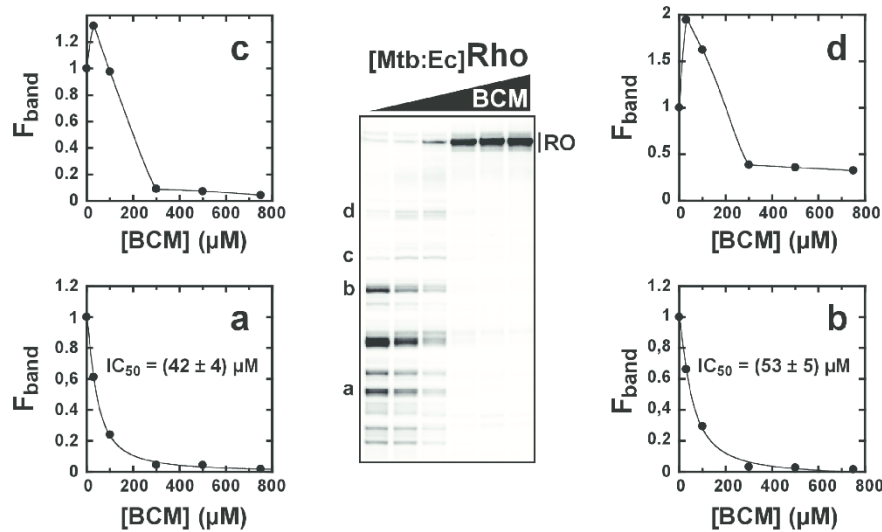
Supplementary figure 3: The MtbRho enzyme is a directional helicase requiring a 5'-ssRNA tail upstream from its duplex target. Unwinding experiments were performed in the presence of ATP under standard conditions (see methods) with substrate D which contains the same 14bp duplex region than substrate C but located upstream rather than downstream from the ssRNA tail (see diagram, inset). Although the duplex in substrate D is slightly less stable than in substrate C, MtbRho does not accelerate duplex unwinding with this substrate (left graph). The same is true for EcRho which even displays a slight 'annealing' activity with substrate D as evidenced by a bell-shaped unwinding curve. By contrast, the non-directional DEAD-box protein Ded1 indiscriminately increases unwinding of the duplexes in substrates C and D. Experiments with Ded1 were performed as described in Ref. 44.



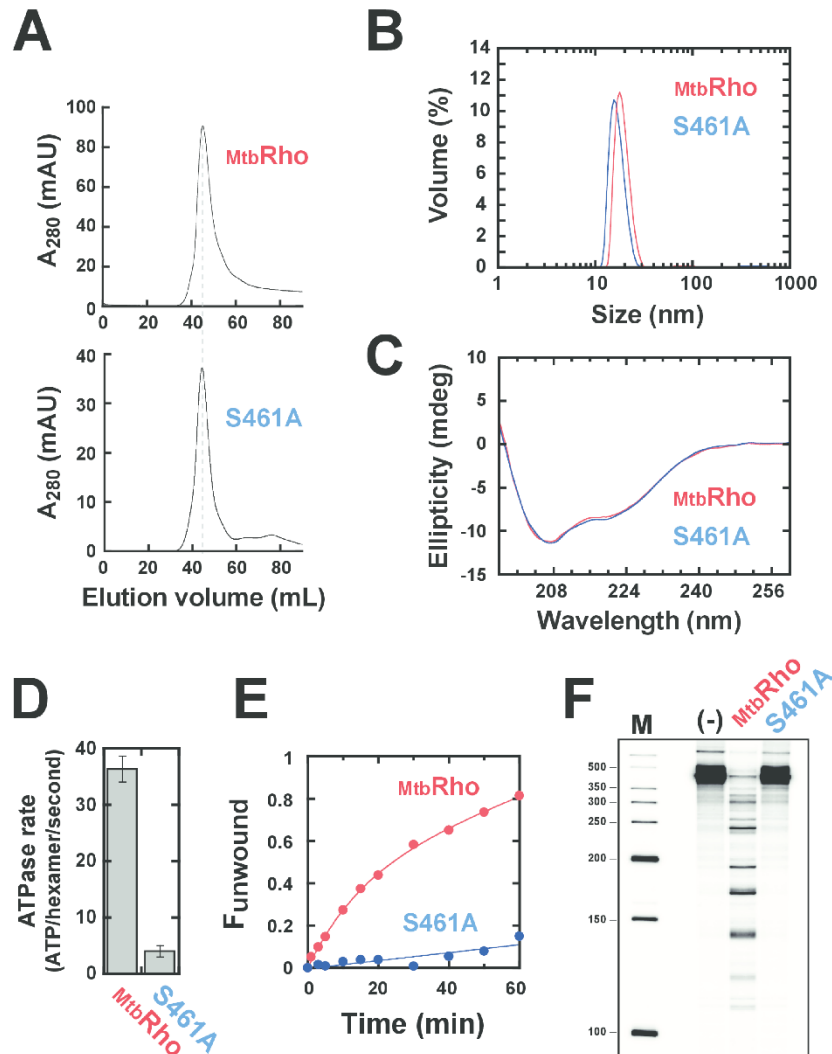
Supplementary figure 4: Single-round transcriptions with bead-affixed transcription complexes containing biotinylated λ tR1 DNA templates. The 32 P-labeled complexes, halted at +24 (by privation of CTP from the initiation mixture), were immobilized and purified on streptavidin-coated magnetic beads. They were then incubated with the Rho variants indicated above gel lanes before being 'chased' with 75 μ M rNTPs. The supernatant (S) and bead (B) fractions were then separated on a Magrack (GE-healthcare) before analysis by denaturing PAGE. The presence of the Rho-dependent transcripts in supernatants confirms that they stem from termination rather than transcriptional pausing or arrest. **(A)** Comparison of the RNA release efficiencies of the various Rho proteins. **(B)** Effect of the presence of BCM (0 or 750 μ M) in reactions performed with the representative WT *Mtb*Rho or the [*Mtb*:Ec]Rho chimera.



Supplementary figure 5: Normalized electropherograms of transcription termination reactions performed under standard 'chase' conditions (as in Figure 6B).



Supplementary figure 6: Upstream shift of the termination window obtained with the $[Mtb:Ec]Rho$ chimera as a function of BCM concentration. Graphs show normalized fractions of products released at representative termination positions as a function of BCM concentration (data averaged from 3 independent experiments).



Supplementary figure 7: Characterization of the single-point mutant S461A of $MtbRho$ by (A) gel filtration on a Sephacryl S-300 HR column, (B) diffusion light scattering, (C) circular dichroism, and (D) steady-state ATPase hydrolysis measurements. The Q-loop mutation greatly alters the capacity of $MtbRho$ to unwind duplex C (E) or to trigger transcription termination (F). Data for WT $MtbRho$ are shown for comparison.

ARTICLE III:
RNA REMODELING BY
BACTERIAL GLOBAL REGULATOR CSRA
PROMOTES RHO-DEPENDENT
TRANSCRIPTION TERMINATION

Genes & Development (2014)

RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination

Nara Figueroa-Bossi,¹ Annie Schwartz,² Benoit Guillemardet,¹ François D'Heygère,² Lionello Bossi,^{1,3} and Marc Boudvillain^{2,3}

¹CNRS UPR3404, Centre de Génétique Moléculaire, Gif-sur-Yvette 91198, France, affilié à l'Université Paris Sud XI, Orsay 91405, France; ²CNRS UPR4301, Centre de Biophysique Moléculaire, Orléans 45071, France, affilié à l'Université d'Orléans, Orléans 45100, France

RNA-binding protein CsrA is a key regulator of a variety of cellular processes in bacteria, including carbon and stationary phase metabolism, biofilm formation, quorum sensing, and virulence gene expression in pathogens. CsrA binds to bipartite sequence elements at or near the ribosome loading site in messenger RNA (mRNA), most often inhibiting translation initiation. Here we describe an alternative novel mechanism through which CsrA achieves negative regulation. We show that CsrA binding to the upstream portion of the 5' untranslated region of *Escherichia coli* *pgaA* mRNA—encoding a polysaccharide adhesin export protein—unfolds a secondary structure that sequesters an entry site for transcription termination factor Rho, resulting in the premature stop of transcription. These findings establish a new paradigm for bacterial gene regulation in which remodeling of the nascent transcript by a regulatory protein promotes Rho-dependent transcription attenuation.

[Keywords: CsrA; Rho factor; transcription attenuation; bacterial gene regulation]

Supplemental material is available for this article.

Received February 17, 2014; revised version accepted April 18, 2014.

Bacterial adaptation to changing environments relies on the ability of the bacterial cell to coordinately regulate groups of genes in response to chemical and/or physical signals. A variety of transcriptional and post-transcriptional mechanisms, often interconnected, underlie these global responses. One such mechanism involves a small (61-amino-acid) protein, named CsrA, acting as a post-transcriptional regulator. Originally discovered as a repressor of glycogen biosynthesis in *Escherichia coli* (Romeo et al. 1993), CsrA was subsequently found to regulate other growth phase-dependent processes such as biofilm formation and quorum sensing in *E. coli* and *Vibrio cholerae* (Jackson et al. 2002; Lenz et al. 2005; Wang et al. 2005; Yakhnin et al. 2011) and host invasion in *Salmonella* (Altier et al. 2000; Lawhon et al. 2003; Martinez et al. 2011). The CsrA structural homologs RsmA/RsmE perform similar functions in *Pseudomonas* and *Erwinia* species, where they participate in the regulation of secondary metabolism (Chatterjee et al. 1995; Cui et al. 1995; Pessi et al. 2001; Lapouge et al. 2007). In all of these systems, CsrA and its homologs act as translational repressors; they interact with messenger RNA (mRNA) se-

quences overlapping ribosome-binding sites and thus block translation initiation by occluding these sites (for reviews, see Babitzke and Romeo 2007; Lapouge et al. 2008; Romeo et al. 2013). Occasionally, CsrA can also act as a positive regulator, protecting mRNA against degradation by RNase E (Yakhnin et al. 2013) or remodeling translation initiation regions in ways that stimulate ribosome binding (Patterson-Fortin et al. 2013; Sterzenbach et al. 2013).

A central aspect in CsrA biology is in the mechanism that regulates CsrA activity. This occurs as a result of the production of two small RNAs, CsrB and CsrC, which contain several CsrA-binding sites and sequester the protein by molecular mimicry (Babitzke and Romeo 2007; Lapouge et al. 2008; Romeo et al. 2013). The *csrB* and *csrC* genes are under the control of the BarA/UvrY (SirA) two-component signal transduction system and are transcriptionally activated in the presence of short chain carboxylic acids and other products of carbon metabolism (Takeuchi et al. 2009; Chavez et al. 2010). In addition, CsrA indirectly activates *csrB* and *csrC* transcription via the BarA/UvrY

³Corresponding authors

E-mail bossi@cgm.cnrs-gif.fr

E-mail marc.boudvillain@cnrs-orleans.fr

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.240192.114>.

© 2014 Figueroa-Bossi et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Figueroa-Bossi et al.

(SirA) system by an unknown mechanism. Combined with the ability of CsrA to feedback-inhibit translation of its own mRNA (Gudapaty et al. 2001; Suzuki et al. 2002), this provides an additional homeostatic loop that likely allows CsrA activity to be finely adjusted at all times (Romeo et al. 2013). Two novel mechanisms modulating CsrA activity/availability were unveiled recently. In one system, McaS, a regulatory small RNA known to function via classical Hfq-dependent RNA pairing, was found to bind CsrA and relieve repression of some CsrA targets (Jorgensen et al. 2013). In another system, not a small RNA, but rather the 5' leader region of fimbrial mRNA cooperates with CsrB and CsrC in titrating CsrA, preventing CsrA-mediated activation of a second hierarchically controlled fimbrial mRNA (Sterzenbach et al. 2013).

CsrA binds short RNA sequences containing a centrally positioned GGA motif that is essential for recognition (Dubey et al. 2005). The high conservation of this motif in Shine-Dalgarno sequences accounts for the predominant role of CsrA as a repressor of translation initiation. Since the functional form of CsrA is a homodimer, tandem binding sites are generally required for a productive interaction (Schubert et al. 2007; Mercante et al. 2009), with the protein bridging GGA sites anywhere between 10 and 63 nucleotides (nt) apart (Mercante et al. 2009). Some regulated genes have multiple CsrA-binding sites in their control regions. This is the case for the *pgaABCD* operon of *E. coli*, which encodes proteins needed for the synthesis and export of a polysaccharide adhesin involved in biofilm formation (Itoh et al. 2008). Work from the Romeo and Babitzke laboratories (Wang et al. 2005) identified six CsrA-binding sites in the 234-nt 5' untranslated region (UTR) of *pgaABCD* mRNA and provided in vitro and in vivo evidence that two distal sites, which cover the Shine-Dalgarno sequence and the initiating AUG, are required for translational repression by CsrA. The roles of the other CsrA-binding sites or the reasons for their redundancy have remained elusive.

We became interested in the regulation of the *pgaABCD* locus while data-mining for links between small RNA activity and Rho-dependent transcription termination. This search, prompted by recent findings in our laboratories (Bossi et al. 2012), was made possible by the publication of a high-resolution map of Rho-dependent termination sites in the *E. coli* genome (Peters et al. 2012). Peters et al. (2012) used global RNA-profiling techniques to compare RNA from bacteria treated with the Rho inhibitor bicyclomycin (BCM) or left untreated and, by aligning the two sets of data, were able to identify transcripts specifically terminated by Rho. One of the "BCM significant transcripts" (BSTs) was from the *pgaA* leader region. We noticed that the putative termination site was near the proposed target sequence for the small regulatory RNA McaS, previously shown to activate *pgaA* expression (Thomason et al. 2012). This raised the intriguing possibility that McaS could act by preventing Rho from binding to the RNA. While our subsequent work invalidated this hypothesis, it led us to discover that CsrA modulates the activity of the terminator. CsrA binding to the upstream portion of the *pgaA* 5' UTR prevents the formation of a secondary structure that

otherwise sequesters the *rut* (rho utilization) sequence. Once exposed, the *rut* sequence is recognized by Rho factor, which terminates transcription. This is an entirely new role for CsrA protein and, to the best of our knowledge, is the first example of a regulatory protein that acts by modulating the accessibility of an RNA sequence to the binding of Rho factor.

Results

A Rho-dependent terminator in the pgaA leader region

The study by Peters et al. (2012) tentatively mapped a Rho termination site at position +55 of the *pgaA* 5' UTR. Upon examining the sequence of the region, we detected the presence of a large C>G-rich "bubble," a hallmark of Rho-dependent terminators (Alifano et al. 1991), spanning the interval between +50 and +130 (Fig. 1A). In particular, the sequence contains two sets of four regularly spaced YC dimers (Fig. 1B), either of which could allow a Rho hexamer to anchor the RNA through four contiguous subunits (Boudvillain et al. 2013). The ensemble would then constitute two superimposed (differently phased) *rut* sites. From their position, one would predict transcription to actually terminate around +200, ~150 nt downstream from the site identified by Peters et al. (2012) (Supplemental Fig. S1). This discrepancy is likely ascribable to the trimming of Rho-terminated transcripts by 3'-5' exonuclease activity in vivo (Peters et al. 2012).

The *rut* sequence lies immediately downstream from the 5'-proximal pair of CsrA-binding sites (Fig. 1B) and includes one of the proposed target sites for the small RNA McaS (see above; Supplemental Fig. S2). Analysis of the entire *pgaA* leader sequence by the Mfold algorithm (Zuker 2003) predicted that the segment between +61 and +107 adopts a discontinuous, "hairpin-like" secondary structure (two double-stranded regions separated by a single-stranded bubble), with the CsrA sites on the left (L) arm of the structure, and the 5' half of the putative *rut* sequence on the right (R) arm (Fig. 1C; Supplemental Fig. S2).

As a first step in this study, we sought to ascertain the presence of a Rho-dependent terminator in the 5' UTR of *pgaA* by in vitro transcription. We prepared a DNA template containing the 5' UTR and the proximal portion of the *pgaA* coding sequence fused to the strong promoter T7A1 (Fig. 2A). Transcription of this DNA template in the presence of purified Rho resulted in the formation of short transcripts (Fig. 2B, "term" bands) consistent with RNA release occurring in the distal portion of the 5' UTR, downstream from the sequence encoding the putative *rut* motif (Fig. 2A; Supplemental Fig. S3). This Rho-mediated transcription termination event, however, was dependent on ionic strength and was nearly abolished in the presence of 150 mM KCl (Fig. 2C, lane 4). The inhibitory effect of high KCl concentrations on Rho-dependent termination is not unprecedented (Zou and Richardson 1991) and may reflect suboptimal terminator features at high ionic strength due to, for instance, stabilization of competing RNA structures. Consistent with this idea, termination in the 5' UTR of *pgaA* was restored in the presence of NusG

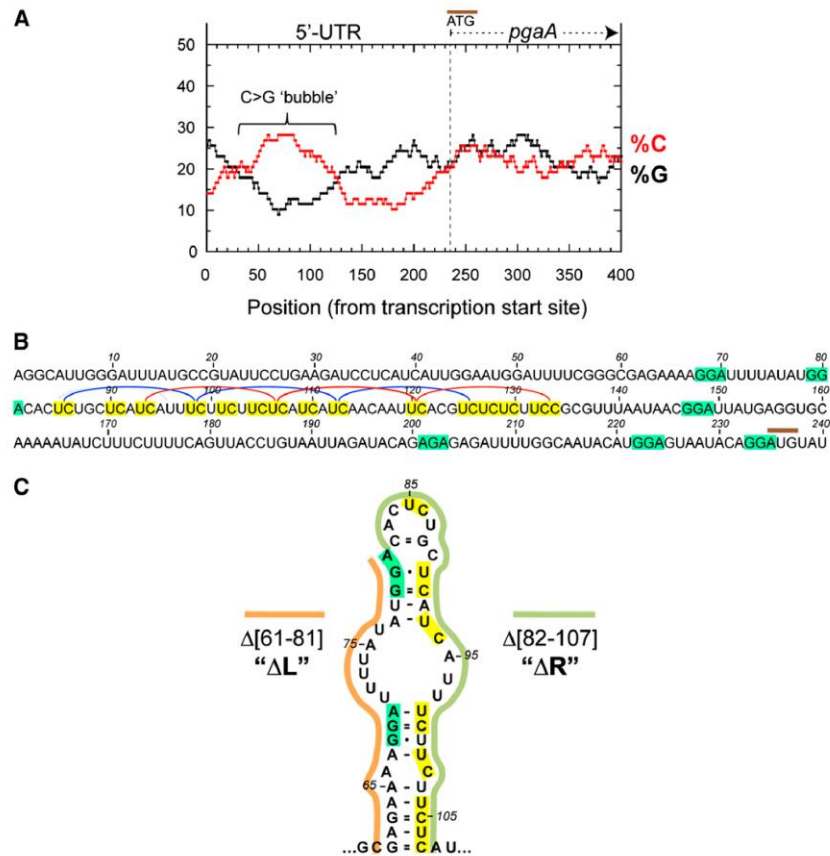


Figure 1. Rho-dependent terminator in the *pgaA* leader region. (A) Graph showing the distributions of cytosine and guanine residues (in 78-nt windows) (Alifano et al. 1991) in the upstream section of the *pgaABCD* operon. (B) Sequence of the *pgaA* leader region. Pyrimidine (YC) dimers (highlighted in yellow) that are “in phase” and could constitute *rut* subsites are linked by blue or red lines. The predicted termination site is ~150 nt downstream from the position identified by Peters et al. (2012), a difference likely ascribable to 3' end trimming of Rho-terminated transcripts by exonucleases in vivo (discussed by Peters et al. 2012). The positions of CsrA-binding sites (core portion highlighted in green) are from Wang et al. (2005). (C) Hairpin-like RNA motif predicted by Mfold in the upstream section of the *pgaA* leader *rut* site (see also Supplemental Fig. S2).

(Fig. 2C, lanes 8–10), a cofactor known to activate termination at suboptimal sites (Burns and Richardson 1995; Peters et al. 2012). CsrA completely relieved the NusG requirement, restoring termination more effectively than NusG itself (Fig. 2C, lanes 5–7). Interestingly, NusG and CsrA appear to act synergistically, inducing termination at promoter-proximal sites (up to position 138) that are not used otherwise (see below; Supplemental Fig. S3). Altogether, these data establish that the *pgaA* leader contains a salt-sensitive, factor-dependent termination signal that may well contribute to the regulation of the *pgaABCD* operon in vivo.

CsrA targets a cis-acting anti-terminator element

We chose to carry out in vivo analyses in *Salmonella*—where we previously constructed a set of congenic strains designed for the study of Rho function (Bossi

et al. 2012)—and optimized the techniques for “scarless” site-directed mutagenesis of chromosomal genes. *Salmonella*’s Rho and NusG proteins are both 99% identical to their *E. coli* counterparts. The Rho allele used in this study—Y80C—also exists in *E. coli*, where it was shown to affect ATP cofactor binding by Rho (Chalissery et al. 2007). The NusG allele—174fs, a –1 frameshift at amino acid position 174—alters the NusG C-terminal sequence (from DFSQVEKA to TSVRLRKRNRS) (Bossi et al. 2012) and presumably affects the ability of NusG to interact with Rho (Mooney et al. 2009).

Since *Salmonella* lacks the *pgaA* operon, the first step of the in vivo study involved moving a DNA fragment comprising the promoter and the 5' UTR of the *E. coli pgaA* gene (from –124 to +234) into the *Salmonella* chromosome. The *pgaA* gene was placed adjacent to a resident *lacZY* operon, with *lacZ*’s initiating AUG at the exact position normally occupied by *pgaA* AUG. We then constructed

Figueroa-Bossi et al.

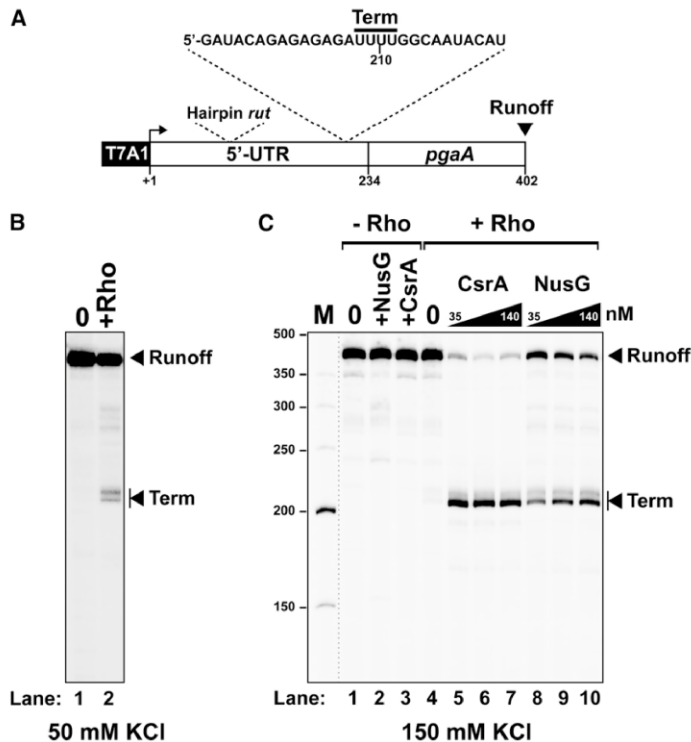


Figure 2. In vitro transcription termination at the Rho-dependent *pgaA* terminator. (A) Schematic representation of the DNA template used for the in vitro transcriptions. Key transcript features are also depicted above the template regions. (B,C) Representative gels show the formation of truncated transcripts resulting from Rho-dependent termination at the *pgaA* site under "low-salt" (50 mM KCl) transcription conditions (B) or after activation by the CsrA and NusG proteins under "high-salt" (150 mM KCl) conditions (C).

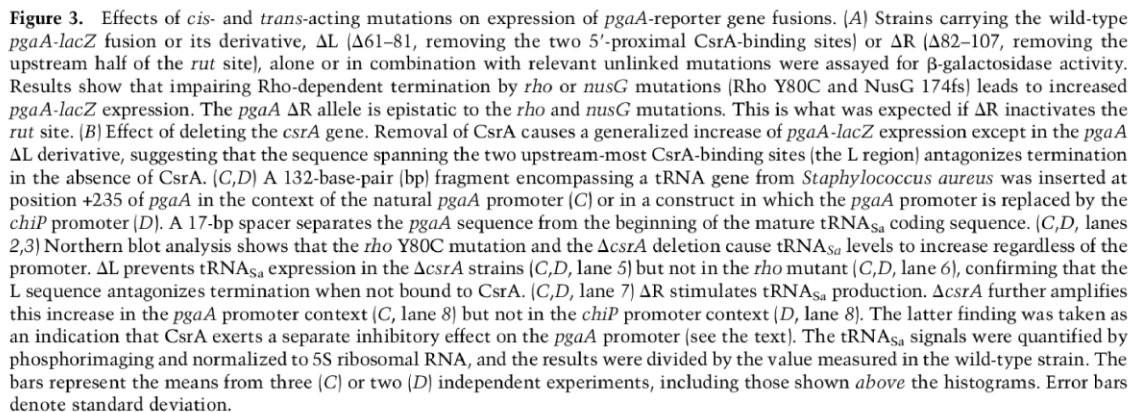
two unmarked deletion derivatives: one lacking the sequence corresponding to the left arm of the hairpin structure (segment 61–81), hereafter referred to as the ΔL deletion, and the other, lacking the right arm (segment 82–107), named ΔR . ΔL removes both 5'-proximal CsrA-binding sites, while ΔR removes approximately the 5' moiety of the *rut* region (Fig. 1B,C), including the presumptive McaS target sequence (Thomason et al. 2012), and was expected to inactivate the termination site. The effects of these alterations on *pgaA-lacZY* expression were analyzed in wild-type and various mutant backgrounds.

As shown in Figure 3A, the ΔR and ΔL deletions have opposite effects on the basal level of expression of the *pgaA-lacZY* fusion. ΔL causes the already low β -galactosidase activity to drop to nearly undetectable levels, whereas ΔR results in an ~10-fold increase. LacZ activity is also increased in the presence of the *rho* and *nusG* mutations (12-fold and twofold, respectively), and this effect is amplified in the ΔL mutant (~60-fold and fivefold). In contrast, the *rho* and *nusG* mutations have no significant effect on *pgaA-lacZY* expression in the ΔR mutant (Fig. 3A). These findings provide in vivo evidence for the involvement of Rho-dependent termination in *pgaA* regulation and tentatively identify the segment missing in the ΔR deletion as the site required for Rho action.

Deleting the *csrA* gene causes a much larger increase in β -galactosidase activity than either the ΔR or *rho* mutations (Fig. 3B, note the different scale). This is not surprising, as

transcripts that escape termination can be expected to remain subjected to translational repression by CsrA (Wang et al. 2005). Furthermore, CsrA down-regulates NhaR, a transcriptional activator of *pgaA* in *E. coli*, by binding to sites within its 5' UTR and inhibiting translation (Pannuri et al. 2012). Although *Salmonella* lacks the *pgaABCD* operon, it carries a close ortholog of the *nhaR* gene. Thus, removal of CsrA will have the additional effect of indirectly activating *pgaA* transcription (see below). Interestingly, the data in Figure 3B show that most of the "activating" effects of the $\Delta csrA$ allele are abolished in the ΔL variant. This epistasis can be interpreted to indicate that the sequence of the left arm of the putative hairpin works as a *cis*-acting anti-terminator element when not bound to CsrA.

In parallel with the above study, we tested our initial hypothesis that McaS might up-regulate *pgaA* by competing with Rho for binding to the *rut* site (see above). Results showed that the ΔR deletion, while removing the presumptive McaS target site (Supplemental Fig. S2), does not abolish the ability of McaS to up-regulate *pgaA-lacZY* (Supplemental Fig. S4). In contrast, up-regulation is lost in a $\Delta L \Delta csrA$ double mutant in spite of the fact that the McaS target sequence is still present in the mRNA (Supplemental Fig. S4). Thus, these data invalidate our hypothesis and suggest that McaS effects depend somehow on the presence of CsrA. In agreement with this conclusion, a report published while this study was under way showed CsrA



fusions was replaced by a heterologous tRNA gene (tRNA_{Sa}) serving as a transcriptional reporter (Bossi et al. 2012). In the new fusion, the distal CsrA-binding site (overlapping the *pgaA*-initiating AUG codon) (Fig. 1B) was eliminated. Northern blot detection of the tRNA_{Sa} confirmed that

To analyze the transcriptional component of *pgaA* regulation separately from CsrA-mediated translational repression, the *lacZ* coding portion of the *pgaA-lacZY*

Figueroa-Bossi et al.

impairing Rho function by mutation or by the ΔR deletion causes a significant increase of readthrough transcription across the *pgaA* leader region, thus corroborating the Rho involvement (Fig. 3C). A sharp increase in tRNA_{Sa} levels is also observed in wild-type and ΔR backgrounds upon removal of the *csrA* gene (Fig. 3C). Two mechanisms could contribute to this increase: (1) relief of CsrA-induced transcription termination and (2) stimulation of *pgaA* transcription initiation by the derepressed NhaR activator (see above; Pannuri et al. 2012). The greater magnitude of the increase in the wild-type construct relative to the ΔR mutant (~20-fold and approximately threefold, respectively) suggests that CsrA-induced transcription termination is the primary cause of the shutoff of tRNA_{Sa} expression. To verify this conclusion, we sought to dissociate the promoter effects from the effects within the leader. To this end, new constructs were made with the *pgaA* leader sequence fused to a different promoter (the *chiP* promoter) (see the Materials and Methods), and the tRNA_{Sa} quantification analysis was repeated with the new strains. As shown in Figure 3D, removal of CsrA still causes tRNA_{Sa} levels to increase in the *pgaA*⁺ strain (cf. lanes 1 and 2). This provides compelling evidence that CsrA represses tRNA_{Sa} synthesis by inducing transcription termination within the *pgaA* leader. The pattern from the $\Delta L \Delta csrA$ double mutant in lane 5 of Figure 3D is consistent with this interpretation. As already observed with the corresponding *lacZ* fusion (Fig. 3A), this construct is locked in an OFF configuration, suggesting that the L arm sequence (Fig. 1C) stimulates readthrough transcription in the absence of CsrA. CsrA binding to this sequence could then be required to counter this effect and thus promote termination.

The data in Figure 3D also show that replacing the *pgaA* promoter with the *chiP* promoter abolishes the further increase of tRNA_{Sa} induced by $\Delta csrA$ in the ΔR strain (Fig. 3, cf. C [lanes 7,8] and D [lanes 7,8]). This confirms that CsrA also affects the activity of the *pgaA* promoter. To verify the NhaR involvement in these effects, an additional construct was made in which the *nhaR* 5' UTR is replaced by the *lacUV5* promoter and the short 5' UTR of *lacZ*. In the NhaR constitutive background, $\Delta csrA$ no longer caused tRNA_{Sa} expression to further increase in the ΔR variant (Supplemental Fig. S5), confirming that this effect results from the relief of *nhaR* repression by CsrA (Pannuri et al. 2012).

Structure-dependent regulation of the *pgaA* terminator

Altogether, the above data point to a key role of RNA structure shown in Figure 1C in the regulation of termination. One alternative possibility was considered in which CsrA regulated translation of a putative leader peptide initiating at nucleotide 78 of the 5' UTR. However, a mutation of the initiating AUG had no effect on *pgaA* expression (Supplemental Fig. S6, see legend for details). The hairpin-like conformation incorporates the 5' half of the *rut* sequence (Fig. 1C). One might expect that RNA folding will limit the accessibility of the sequence to Rho factor. CsrA binding to the L arm of the hairpin would expose the *rut* site and make it fully available for termination. This model

predicts that L arm mutations in the CsrA-binding sites will have opposite effects on *pgaA* regulation depending on whether they increase or decrease the stability of the structure. We proceeded to test the prediction by constructing double- and triple-base replacement mutants (depicted in Fig. 4A). These changes were introduced directly in the chromosome of a strain carrying the *pgaA-lacZY* fusion. The resulting strains were characterized by β -galactosidase assays (described below) and served as a source of template DNA for in vitro analyses. As part of the latter work, RNA transcripts from the upstream portion of the *pgaA* leader (interval +1 to +137) were used to measure CsrA-binding affinity by gel shift assay. This analysis offered an unforeseen yet convincing confirmation for the existence of a postulated secondary structure. We found that some of the *pgaA* leader mutations markedly affected RNA mobility in native gels (Fig. 4B). In particular, mutations predicted to destabilize the hairpin caused a sharp decrease in mobility. The effects could be readily rationalized considering that folding increases the compactness of the molecule, allowing it to migrate faster in the gel. Overall, the migration patterns (Fig. 4B) correlate remarkably well with the structural and free energy predictions from the Mfold program, suggesting that RNA mobility can be used to infer the presence/absence of such structure. For example, the data in Figure 4B indicate that under the conditions of the experiment, wild-type RNA adopts a folded configuration in spite of the imperfect nature of the hairpin structure.

The gel shift experiments provided insight as to the primary and secondary structural elements that contribute to CsrA binding. This analysis showed that the *d1* variant binds CsrA better than wild-type RNA in spite of being altered in the second CsrA-binding site (Fig. 4C). Apparently, reducing the stability of the upper portion of the hairpin structure is more important for CsrA binding than full conformity to the consensus sequence. Consistent with this conclusion, variant *s3*, which is unaffected in the CsrA-binding sites but is predicted to form a more stable hairpin, bound CsrA poorly. In contrast, primary sequence determinants appear critical in the bottom portion of the structure, as inferred from the finding that mutation *d2* nearly completely abolishes binding (Fig. 4C). Combined with the greater effect of mutation *s2* relative to mutation *s1* (note also dominance in the double mutant), these findings suggest that the CsrA–RNA interaction nucleates at the first CsrA-binding site in the bottom portion of the hairpin.

The effects of these changes on the in vivo regulation of the *pgaA-lacZY* fusion were quantified by measuring β -galactosidase activity. This analysis showed that the destabilizing changes lower *pgaA-lacZ* expression in both the *csrA*⁺ (Fig. 4D) and $\Delta csrA$ (Fig. 4E) backgrounds, as predicted if they relieve the CsrA requirement for termination. In contrast, changes that tighten the double-stranded regions of the hairpin show the opposite trend, increasing readthrough transcription. Finding that mutant *s3* is partially deattenuated in the *csrA*⁺ strain (Fig. 4D) despite its sequence changes lying outside the CsrA recognition sequence is consistent with the behavior of this mutant in the gel shift assay and reiterates the

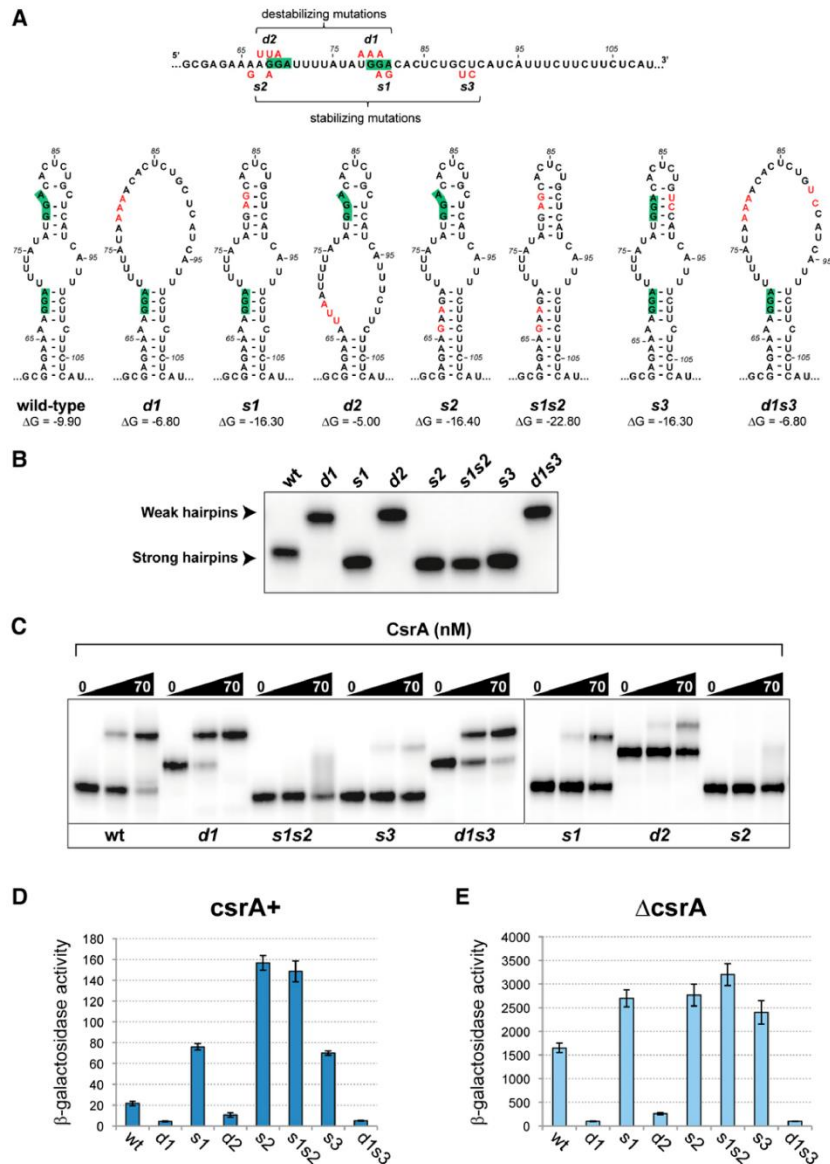


Figure 4. *pgaA* RNA hairpin mutants and their CsrA-binding properties. (A) Diagram showing the positions of the double- and triple-base changes introduced near the CsrA-binding sites in the upstream portion of the *pgaA* 5' UTR and the predicted effects of these changes on the RNA secondary structure. Drawings reproduce the outputs from the Mfold program [Zuker 2003]. (B) Effect of hairpin mutations on RNA mobility in native gels. Changes predicted to increase RNA folding in *A* cause RNA to migrate faster in the gel, while changes predicted to decrease folding cause reduced electrophoretic mobility. (C) Gel retardation assay with purified CsrA. The sequence changes in the bottom portion of the RNA hairpin lower CsrA-binding affinity regardless of their effects on secondary structure. In contrast, the behavior of mutant *d1*, which binds CsrA better than wild-type RNA, suggests that secondary structure determinants are more important than primary structure determinants in the upper portion of the RNA hairpin. (D,E) Effect of RNA conformational changes on *pgaA-lacZ* expression. Strains carrying the *lacZ* fusion derivatives of *pgaA* hairpin mutants were assayed for β -galactosidase activity in *csrA*⁺ (D) and Δ *csrA* (E) backgrounds. Results show that mutations destabilizing the RNA structure inhibit *lacZ* expression regardless of the *csrA* allele. Apparently, these changes improve the accessibility of the *rut* site, rendering CsrA action dispensable. In contrast, stabilizing mutations promote anti-termination. The *s2* and *s1s2* variants show the highest levels of readthrough transcription in *csrA*⁺ but not in Δ *csrA*. This can be rationalized considering that *s2* and *s1s2* abolish CsrA binding completely, whereas all other variants have residual CsrA-binding affinity.

Figueroa-Bossi et al.

importance of single strandedness for CsrA binding. Also consistent with the reduced CsrA binding to the *s2* versus the *s1* leader (Fig. 4C) is the higher level of readthrough transcription in mutant *s2* relative to *s1* (Fig. 4D), which again suggests that the CsrA interaction with the bottom part of the hairpin initiates melting of the structure. Altogether, the results from this analysis suggest that the hairpin-like secondary structure formed by the 61- to 107-nt sequence is programmed to have borderline stability to facilitate the CsrA interaction and the rapid switch to the open conformation following CsrA binding.

To complement the above study, a representative from either side of the stability spectrum (*d1* and *s1s2*) was analyzed in the *in vitro* transcription system. The DNA templates that were used cover the entirety of the *pgaA* 5' UTR and extend a further 100 base pairs (bp) into the *lacZ* gene (Fig. 5A). Experiments performed under high-salt conditions (150 mM KCl) (Fig. 5B) confirmed that CsrA is needed to activate Rho-dependent termination at the wild-type *pgaA* sequence (Fig. 5B, cf. lanes 2 and 3). Remarkably consistent with the *in vivo* data, the CsrA requirement is completely relieved in the *d1* template (Fig. 5B, lanes 6,7), whereas CsrA loses the capability to activate Rho-dependent termination in the *s1s2* template (Fig. 5B, lanes 10,11). Similar observations were made under "low-salt" transcription conditions (50 mM KCl), with the predicted exception that CsrA was no longer needed to activate termination with the wild-type construct (Supplemental Fig. S7). These results confirm that the degree of stability of the hairpin motif ($\Delta G_{WT} = 9.9$ kcal/mol) is a critical factor for the control of the Rho-dependent *pgaA* terminator by CsrA.

Intriguingly, *in vitro* activation of the *pgaA* terminator by NusG appears to be affected by hairpin stability as well (Fig. 5C, cf. lanes 3, 6, and 9). NusG only slightly activates termination at the main sites with the "strong" *s1s2* hairpin construct (Fig. 5C, lane 9) while strongly activating termination at both the main sites and promoter-proximal sites with the "weak" *d1* hairpin construct (Fig. 5C, lane 6, note that band signals for short transcripts are inherently weaker due to fewer internal incorporations of [α - 32 P]-U residues). Again, the wild-type template shows an intermediate behavior, with both CsrA and NusG needed to activate promoter-proximal termination sites (Fig. 5B, lane 4). These data suggest that NusG somehow allows Rho to bypass the requirement for a fully single-stranded *rut* sequence even though the combined actions of both CsrA and NusG produce maximal termination efficiency (Fig. 5, cf. B [lanes 4,8,12] and C [lanes 3,6, 9]). The NusG involvement is consistent with the *in vivo* data showing a moderate but reproducible increase of transcriptional readthrough at the *pgaA* 5' UTR terminator in a *nusG* mutant (Fig. 3A,B) and with the significant overlap in the 3' end patterns of *pgaA* transcripts obtained from BCM-treated wild-type *E. coli* and a *nusG* mutant, as detected on tiling arrays (Peters et al. 2012).

Discussion

The classical mechanism for regulating gene expression through the activity of transcription termination factor

Rho involves the action of proteins that render the transcription elongation complex refractory to termination (anti-termination). Thus, the classical scenario (e.g., phage λ 's N protein) has termination as the default setting and anti-termination as the active step; furthermore, the regulatory target is RNA polymerase rather than Rho itself. The work described here uncovered an alternative mechanism for Rho-mediated gene regulation. In this mechanism, the active step consists of inducing termination at a *rut* site that is normally held invisible to Rho because it is part of a secondary structure. Termination results from the action of a protein, CsrA, that, upon binding to the 5' arm of the structure, exposes the *rut* sequence, making it available for Rho binding (Fig. 6). The mechanism relies on the exquisite fine-tuning of an RNA secondary structure: stable enough to hold the *rut* site hidden from Rho but sufficiently unstable to allow nucleation of the CsrA-RNA complex and the opening of the structure following CsrA binding. Also important to consider is that in order to induce Rho-dependent termination, CsrA must be able to bind the RNA while the latter is being synthesized and before RNA polymerase escapes the termination "window." Presumably, the fact that the secondary structure has not yet formed at this stage can favor the interaction. This might suggest that the 5'-proximal CsrA-binding site (the base of the hairpin stem) is the "seed" sequence. The greater effect of changes in this region on termination and CsrA binding further supports this conclusion. The extent to which other features in the region might contribute to CsrA binding and Rho-dependent termination remains to be explored. Elongation rates and interaction with RNA-binding proteins are viewed as major factors affecting cotranscriptional RNA folding (Pan and Sosnick 2006). The important role played in this study by elongation factor NusG—a protein thought to inhibit transcriptional pausing (Artsimovitch and Landick 2000) and have some RNA-binding capability (Steiner et al. 2002)—further supports the idea that cotranscriptional folding is a key component in the functioning of the *pgaA* attenuator structure.

Besides inducing Rho-dependent termination in the *pgaA* leader region, CsrA represses *pgaA* mRNA translation by binding to a pair of sites near the initiating AUG codon (Wang et al. 2005). The sequence layout is such that translational repression can operate only on transcripts that escape CsrA-induced termination. This raises the question of the biological role of such a redundancy, particularly since our data suggest that *pgaA* attenuation is sufficient to completely shut off *pgaA* mRNA synthesis. Perhaps the two forms of regulation target separate classes of molecules: attenuation mainly directed to nascent RNAs, and translational repression acting on transcripts that are already made at the time of the regulatory switching. This would ensure the rapidity and efficacy of *pgaA* silencing under conditions prompting bacteria to switch from the sessile to the planktonic lifestyle. Alternatively, the two-level regulation might be designed to do just the opposite; that is, make the transition more gradual. RNA footprint data from the initial analysis of the CsrA:*pgaA* mRNA interaction (Wang et al. 2005) suggest that CsrA binds the translation regulatory sites

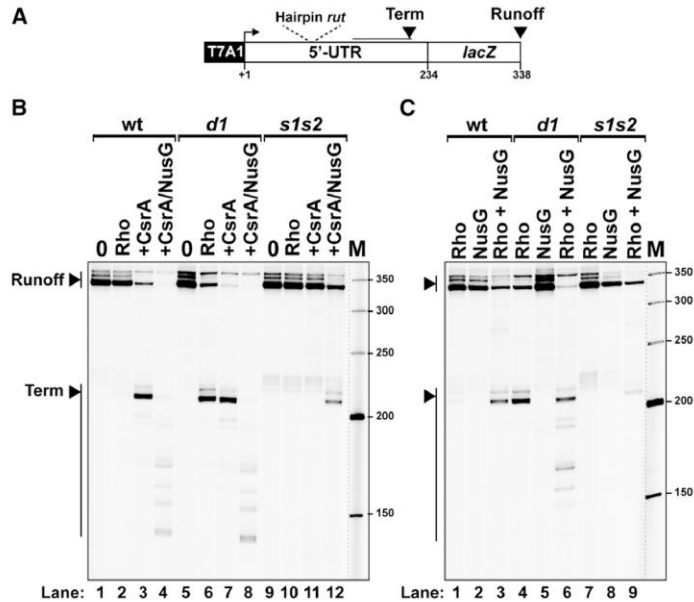


Figure 5. Effect of representative *pgaA* hairpin mutations (*d1* and *s1s2*) on the functioning of the Rho termination site in vitro. (A) Schematic representation of the DNA template used. (B,C) Representative gels show the activation of the terminator variants by CsrA (B) or NusG (C). Results show that *pgaA* hairpin mutation *d1*, expected to favor melting of the secondary structure (Fig. 4B), relieves the requirement of CsrA for in vitro termination by Rho. In contrast, the *s1s2* variant abolishes termination, and CsrA cannot restore it. Thus, these data closely parallel the results from the in vivo analysis (Fig. 4D). B (lanes 4,8,12) also shows that CsrA and NusG act synergistically to stimulate Rho-dependent termination.

with greater affinity than the 5'-proximal sites. Thus, one might envision that *pgaA* translational repression would start early on during a transition and that complete shutoff will occur only after CsrA has sufficiently accumulated.

A further possibility is that the double regulatory layer could have evolved to allow independent sensing of different environmental cues. For example, one could imagine that some physical stimuli or ligands could directly influence the ability of the *pgaA* terminator to switch from the closed to the open configuration so as to induce termination under conditions in which CsrA is still limiting. In other words, the *pgaA* terminator could have the properties of a bona fide riboswitch. A precedent for this type of dual regulation was recently set by the demonstration that the molybdenum cofactor-sensing riboswitch of the *moaA* gene of *E. coli* includes binding sites for CsrA, which activates *moaA* expression post-transcriptionally (Patterson-Fortin et al. 2013). Other recent findings have linked riboswitch function to Rho activity. In the *mgtA* and *ribB* genes, RNA conformational changes induced by the binding of Mg^{2+} and flavin mononucleotide, respectively, modulate the activity of Rho-dependent transcription terminators upstream of the structural portions of the genes (Hollands et al. 2012). A "conditional" Rho-dependent terminator is also found in the *chiPQ* operon; here, however, the terminator is embedded within a translated region and becomes active following the inhibition of *chiP* mRNA translation by a regulatory small RNA (Bossi et al. 2012). The present study provides the first example of a Rho-dependent terminator activated by a protein-induced RNA conformational rearrangement.

One might predict that other genes in the CsrA network will be found to be regulated by a mechanism similar to that described here. A survey of transcriptomic data in the literature reveals some overlaps between the lists of RNAs

copurified with CsrA in *E. coli* (Edwards et al. 2011) and transcripts found extended at their 3' ends in BCM-treated cells (Peters et al. 2012). However, our inspection and Mfold analysis of the sequences in the 5' regions of the shared RNAs failed to detect features reminiscent of the *pgaA* hairpin motif described here. The possibility remains that CsrA and Rho regulatory circuits are interconnected in a more elaborate way not immediately apparent through the above analysis. Moreover, it should be noted that *pgaA* is absent from the list of CsrA-bound transcripts (Edwards et al. 2011), suggesting that the list still has to be completed. Finally, the general nature and robustness of the attenuation mechanism described here suggest that structural remodeling of Rho-dependent signals may well occur in response to RNA binding by other regulatory proteins.

In terms of regulatory responses, transcription termination and anti-termination can be equated to repression and activation of initiation, respectively. A theory of gene regulation elaborated by Savageau in the late 1970s (Savageau 1977) and 1980s (Savageau 1989) posits that genes for which a high level of expression is most frequently required (high demand) will tend to be controlled in a positive manner (by activators or anti-terminators), whereas genes whose products are only required in specific instances (low demand) will be mainly controlled by negative regulatory elements (repressors or proterminators) (Savageau 1977). The theory interpreted classical patterns of gene regulation in *E. coli*, phage λ , and yeast. However, the lack of examples of regulatory proteins acting as transcription proterminators left the picture somewhat incomplete (Savageau 1989). The demonstration, presented here, that CsrA can perform such function allows completing the symmetry. Furthermore, since the requirement

Figuroa-Bossi et al.

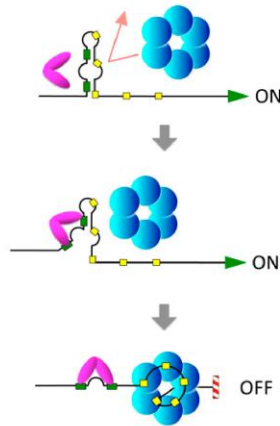


Figure 6. Model for CsrA-induced Rho-dependent transcription attenuation in the *pgaA* leader region. The initial portion of the *pgaA* 5' UTR folds into an imperfect hairpin-like structure that incorporates two CsrA-binding sites in the left arm (green boxes) and part of a Rho-binding site (*rut*) in the right arm (row of yellow boxes). The structure renders the *rut* site inaccessible to Rho but is sufficiently unstable to allow nucleation of a CsrA:RNA complex (presumably starting at the base of the hairpin). Thus, when CsrA is limiting, transcription reads through the terminator and proceeds into *pgaABCD* structural genes, whereas, in the presence of CsrA, formation of the CsrA:RNA complex unfolds the structure, exposing the *rut* site and causing premature transcription termination.

for the *pgaABCD* gene products may well represent a relatively infrequent ("low demand") event in the lifestyle of most *E. coli* strains, our findings lend further support for Savageau's model (Savageau 1989).

Materials and methods

Bacterial strains and culture conditions

Strains used in this study were all derived from *Salmonella enterica* serovar Typhimurium strain MA3409, a derivative of strain LT2 cured for the Gifsy-1 prophage (Figuroa-Bossi et al. 1997). The genotypes of strains used are listed in Supplemental Table S1. Bacteria were cultured at 37°C in liquid LB broth (Bertani 2004) or LB broth solidified by the addition of 1.5% Difco agar. Antibiotics (Sigma-Aldrich) were included at the following final concentrations: 10 µg/mL chloramphenicol, 50 µg/mL kanamycin monosulphate, 100 µg/mL sodium ampicillin, 80 µg/mL spectinomycin dihydrochloride, and 25 µg/mL tetracycline hydrochloride. LB plates containing 40 µg/mL 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-gal; from Sigma) or MacConkey agar plates containing 1% lactose (Macconkey 1905) were used to monitor *lacZ* expression in bacterial colonies. Liquid cultures were grown in New Brunswick gyratory shakers, and growth was monitored by measuring the optical density at 600 nm with a Shimadzu UV mini 1240 spectrophotometer.

Enzymes and chemicals

T4 polynucleotide kinase, T4 DNA ligase, Taq DNA polymerase, and restriction enzymes were purchased from New England

Biolabs. Pfu-Turbo DNA polymerase was from Stratagene. Sigma-saturated RNA polymerase from *E. coli* was obtained from Epicentre. DNA oligonucleotides were custom-synthesized by Sigma-Aldrich, Eurogentec, or Eurofins MWG/Operon. Most chemicals were purchased from Sigma-Aldrich. Acrylamide-bisacrylamide and other electrophoresis reagents were from Bio-Rad and Amresco. Agarose was from Invitrogen. Hybond-N⁺ membranes and hybridization buffer used for Northern blot analysis were from GE Healthcare and Applied Biosystems-Ambion, respectively. The rNTPs were from Promega, and the ³²P-NTPs were from PerkinElmer or Hartmann Analytic. ³²P-labeled nucleic acids were detected by phosphorimaging using ImageQuant software.

Genetic techniques

Generalized transduction was performed using the high-frequency transducing mutant of phage P22, HT 105/1 *int-201* (Schmiegier 1972), as described (Lemire et al. 2011). Chromosomal engineering (recombineering) was carried out by the λ *red* recombination method (Datsenko and Wanner 2000; Murphy et al. 2000; Yu et al. 2000), implemented as in Datsenko and Wanner (2000). "Scarless" recombineering was performed as described (Bossi et al. 2012). Donor DNA fragments were generated by PCR using plasmid DNA, chromosomal DNA, or DNA oligonucleotides as templates. Amplified fragments were electroporated into appropriate strains harboring the conditionally replicating plasmid pKD46, which carries the λ *red* operon under the control of the P^{BAD} promoter (Datsenko and Wanner 2000). Bacteria carrying pKD46 were grown at 30°C in the presence of ampicillin and exposed to 10 mM arabinose for 3 h prior to preparation of electrocompetent cells. Electroporation was carried out using a Bio-Rad MicroPulser under the conditions specified by the manufacturer. Recombinant colonies were selected on LB plates containing the appropriate antibiotic. Constructs were verified by PCR and DNA sequence analysis (performed by GATC Company). The complete list of primers used in this study are in Supplemental Table S2, and details on the construction of several alleles are in Supplemental Table S3.

Construction of relevant alleles

pgaA-lacZY Two constructs were made that convert a chromosomal *chiP-lacZY* gene fusion (Figuroa-Bossi et al. 2009) into a *pgaA-lacYZ* fusion. In the first construct, a DNA fragment spanning the portion of the *pgaABCD* operon from -124 to +234 (amplified from *E. coli* strain MG1655 with primers ppL56 and ppL59) (Supplemental Table S2) was used to replace the region between -253 and +89 of the *chiP* gene (spanning the transcription regulatory region and the 5' UTR) (Plumbridge et al. 2014). This yielded strain MA11469. In the second construct, the exchanged segment covers only the 5' UTR of *chiP* (from +1 to +89), replaced by the *pgaA* segment from +1 to +234 (amplified with primers ppL57 and ppL59); thus, the fusion is transcribed from the *chiP* promoter (strain MA11470). In both constructs, the *lacZ*-initiating AUG falls at the exact same position normally occupied by *pgaA* AUG. Derivatives of MA11469 and/or MA11470 used in this study include strains carrying (1) tRNA_{Ser} gene fusions in place of *lacZ*, (2) deletions in relevant portions of the *pgaA* 5' UTR (ΔL and ΔR), and (3) point mutations affecting the hairpin structure in the 5' UTR (see below).

pgaA hairpin mutants Short DNA fragments (typically ~100 bp) containing the desired mutations were obtained by annealing suitably designed, partially complementary DNA oligonucleotides and filling in the single-stranded regions by PCR under standard

amplification conditions ("reciprocal priming"). DNA fragments were introduced into a strain MA11469 derivative, a *tetRA* insertion in the *pgaA* leader (Supplemental Tables S1, S3), and tetracycline-sensitive recombinants were selected for resistance to fusaric acid [Bochner et al. 1980].

$\Delta csrA::cat$ A deletion removing nearly the entirety of the *Salmonella csrA* coding sequence (167 out of 183 bp) was obtained by replacing this region with the FRT-*cat*-FRT module of plasmid pKD3 [Datsenko and Wanner 2000], amplified with primers ppM17 and ppM18 (Supplemental Table S2). As previously reported [Altier et al. 2000], we found the *csrA* deletion to severely affect growth. A deletion of *glgCAP* operon, reported to suppress this growth defect [Timmermans and Van Melder 2009], failed to restore normal growth. However, suppressors of the growth defect arose spontaneously at high frequency [Altier et al. 2000]. Introduction of the *pgaA-lacZ* fusion in four independent fast-growing isolates showed all four of them to express high levels of β -galactosidase activity. However, when the $\Delta csrA::cat$ allele was replaced by wild-type *csrA*, *pgaA-lacZ* expression returned to normally repressed levels in all four strains, indicating that the defect in *pgaA* regulation is entirely ascribable to the loss of CsrA activity. One of the $\Delta csrA::cat$ *sup* strains (MA11529) was used throughout this study.

$\Delta[nhaA-nhaR']::aph P^{lac}-nhaR$ A DNA fragment carrying the *aph* (KanR) gene and a *lacUV5* promoter was amplified from plasmid pNFB20, with primers ppP23 and ppP24 (Supplemental Table S2) used to replace the region encompassing the *nhaA* gene and the 5' UTR of the adjacent *nhaR* gene in the *Salmonella* chromosome by λ *red* recombination (Supplemental Table S3). In the strain obtained (MA11914), the *nhaR* gene is expressed from the *lacUV5* promoter and is no longer sensitive to translational repression by CsrA [Pannuri et al. 2012]. Plasmid pNFB20 is an oriR6K plasmid carrying the *aph* gene and *lacUV5* side by side in parallel orientation.

Construction of plasmid pBR-*mcaS*

Recombinant plasmid pBR-*mcaS*, which carries the *mcaS* gene under the control of the *lac* promoter, was constructed by cloning a DNA fragment amplified from *E. coli* strain MG1655 (with primers ppL67 and ppL68) into plasmid pBR-plac [Guillier and Gottesman 2006] as described by Thomason et al. [2012].

Measurement of β -galactosidase activity

β -Galactosidase activity was assayed in toluene-permeabilized cells as described [Miller 1992] and is expressed in Miller units throughout this study. Typically, measurements were performed on duplicate or triplicate cultures grown in late exponential phase ($OD_{600} \approx 0.7$). All experiments included parental or reference strains as normalization controls. Standard deviations were generally <5% of the mean.

RNA extraction and Northern analysis

RNA was prepared by the acid hot phenol method from exponentially growing cells (OD_{600} of 0.35) as previously described [Bossi and Figueroa-Bossi 2007]. RNA was fractionated on an 8% polyacrylamide/8 M urea gel and transferred to Hybond-N⁺ membranes. Blots were hybridized to DNA oligonucleotides ppJ62 and ppB10, complementary to *Staphylococcus aureus* tRNA^{Arg} and *Salmonella* 5S ribosomal RNA, respectively (Supplemental Table S2), labeled at the 5' ends with T4 polynucleotide kinase. Hybridization patterns were analyzed by phos-

phorimaging, and RNA bands were quantified using the Image-Quant program.

Preparation of proteins

Plasmid pEt28b-CsrA, encoding CsrA fused to a C-terminal His6 tag, was obtained by inserting phosphorylated oligonucleotides CsrA1 to CsrA3 [annealed to complementary oligonucleotides CsrA4 to CsrA6] (Supplemental Table S2) between the NcoI and XhoI sites of plasmid pET28b (EMD Millipore) in a "one-pot" ligation reaction. CsrA protein was overexpressed in BL21(DE3);pEt28b-CsrA and purified by affinity chromatography on HisPur cobalt resin (Thermo Scientific) following a protocol similar to that previously described [Liu and Romeo 1997]. Recombinant Rho and NusG proteins were overexpressed and purified as described previously [Rabhi et al. 2011]. Concentrations of active oligomeric forms (Rho hexamers, CsrA dimers, and NusG monomers) are indicated throughout the article.

In vitro transcription termination experiments

DNA templates were prepared by standard PCR amplification of genomic DNA using forward primer *pgaA*1, which contains the sequence of the T7A1 promoter fused to the 31-nt upstream segment of the *pgaA* leader sequence, and reverse primer *pgaA*2 (Fig. 2A) or *pgaA*6 (Fig. 5A). Standard transcription termination experiments were performed as described previously [Rabhi et al. 2011], with minor modifications. Briefly, 18- μ L mixtures containing the DNA templates (5 nM final concentration), 20 nM RNA polymerase, 0 or 70 nM Rho, 0.2 U/ μ L SUPERase-In (Ambion), 0 or 70 nM CsrA (unless indicated otherwise), and 0 or 140 nM NusG (unless indicated otherwise) in transcription buffer (40 mM Tris-HCl at pH 8.0, 50 or 150 mM KCl, 5 mM MgCl₂, 1.5 mM DTT) were preincubated for 10 min at 37°C. Next, 2 μ L of transcription initiation mix [2 mM ATP, GTP, and CTP; 0.2 mM UTP; 2.5 μ Ci/ μ L [α -³²P]-UTP; 250 μ g/mL rifampicin in transcription buffer] was added before incubation for 20 min at 37°C. Transcription reactions were stopped with 4 μ L of 0.5 M EDTA and 2 μ L of 0.25 mg/mL tRNA before extraction with a 5:24:1 phenol:chloroform:isoamyl alcohol mix and precipitation with ethanol. Reaction pellets were dissolved in denaturing loading buffer (95% formamide, 15 mM EDTA) and analyzed by denaturing 6% polyacrylamide gel electrophoresis. RNA release at the *pgaA* terminator was also monitored in single-round "chase" transcription experiments with bead-affixed transcription elongation complexes, as detailed in the Supplemental Material.

Electrophoretic mobility assays

RNA transcripts corresponding to the 1–137 region of the *pgaA* leader (including wild-type or mutant hairpin variants) were prepared by standard in vitro transcription with T7 RNA polymerase and PCR-generated DNA templates. Transcripts were purified by denaturing PAGE and stored at –20°C in ME buffer (10 mM MOPS at pH 6.5, 1 mM EDTA). Transcripts were dephosphorylated in small batches (~10 pmol) with alkaline phosphatase (Roche Applied Science) before being labeled with [³²P]-ATP and T4 polynucleotide kinase and subsequently purified by 7% denaturing PAGE. Concentrations of transcripts were determined from their absorptions at 260 nm using a Nanodrop spectrophotometer. For electrophoretic mobility assays, the ³²P-labeled transcripts (0.1 nM final concentration) were incubated for 20 min at 30°C in EMSA buffer (40 mM Tris-HCl at pH 8.0, 150 mM KCl, 5 mM MgCl₂, 0.2 mM DTT, 50 μ g/mL BSA, 30 μ g/mL tRNA, 2% glycerol) in the presence or absence of CsrA protein (final concentrations are

Figueroa-Bossi et al.

indicated in the figures). The mixtures were then mixed with Ficoll-400 (4%, final concentration) and analyzed by 6% native PAGE. Gels were run at 10 V/cm for 5 h at 19°C.

Acknowledgments

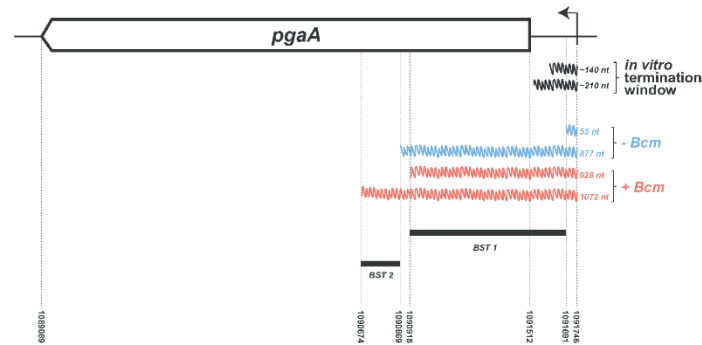
We thank Damien Le Minoux for help in strain construction during the late stages of this work, and Emilie Soares for the preparation of materials used in *in vitro* experiments. We are grateful to Maude Guillier (IBPC, CNRS, Paris) for the gift of plasmid pBR-*plac*. This work was supported by CNRS core funding and grants from the French Agence Nationale de la Recherche (ANR-13-BSV3-0005-01 to N.E.-B., and ANR-13-BSV3-0005-02 to M.B.).

References

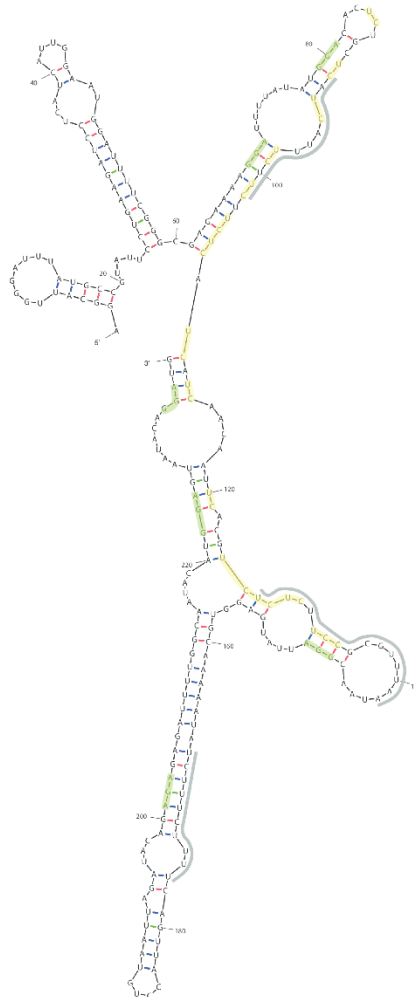
- Alifano P, Rivellini F, Limauro D, Bruni CB, Carlomagno MS. 1991. A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* **64**: 553–563.
- Altier C, Suyemoto M, Lawhon SD. 2000. Regulation of *Salmonella enterica* serovar Typhimurium invasion genes by *csrA*. *Infect Immun* **68**: 6790–6797.
- Artsimovitch I, Landick R. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci* **97**: 7090–7095.
- Babitzke P, Romeo T. 2007. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* **10**: 156–163.
- Bertani G. 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol* **186**: 595–600.
- Bochner BR, Huang HC, Schieven GL, Ames BN. 1980. Positive selection for loss of tetracycline resistance. *J Bacteriol* **143**: 926–933.
- Bossi L, Figueroa-Bossi N. 2007. A small RNA downregulates LamB maltoporin in *Salmonella*. *Mol Microbiol* **65**: 799–810.
- Bossi L, Schwartz A, Guillemardet B, Boudvillain M, Figueroa-Bossi N. 2012. A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes Dev* **26**: 1864–1873.
- Boudvillain M, Figueroa-Bossi N, Bossi L. 2013. Terminator still moving forward: expanding roles for Rho factor. *Curr Opin Microbiol* **16**: 118–124.
- Burns CM, Richardson JP. 1995. NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator. *Proc Natl Acad Sci* **92**: 4738–4742.
- Chalissery J, Banerjee S, Bandey I, Sen R. 2007. Transcription termination defective mutants of Rho: role of different functions of Rho in releasing RNA from the elongation complex. *J Mol Biol* **371**: 855–872.
- Chatterjee A, Cui Y, Liu Y, Dumenyo CK, Chatterjee AK. 1995. Inactivation of *rsmA* leads to overproduction of extracellular pectinases, cellulases, and proteases in *Erwinia carotovora* subsp. *carotovora* in the absence of the starvation/cell density-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone. *Appl Environ Microbiol* **61**: 1959–1967.
- Chavez RG, Alvarez AF, Romeo T, Georgellis D. 2010. The physiological stimulus for the BarA sensor kinase. *J Bacteriol* **192**: 2009–2012.
- Cui Y, Chatterjee A, Liu Y, Dumenyo CK, Chatterjee AK. 1995. Identification of a global repressor gene, *rsmA*, of *Erwinia carotovora* subsp. *carotovora* that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting *Erwinia* spp. *J Bacteriol* **177**: 5108–5115.
- Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci* **97**: 6640–6645.
- Dubey AK, Baker CS, Romeo T, Babitzke P. 2005. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* **11**: 1579–1587.
- Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, et al. 2011. Circuitry linking the Csr and stringent response global regulatory systems. *Mol Microbiol* **80**: 1561–1580.
- Figueroa-Bossi N, Coissac E, Netter P, Bossi L. 1997. Unsuspected prophage-like elements in *Salmonella typhimurium*. *Mol Microbiol* **25**: 161–173.
- Figueroa-Bossi N, Valentini M, Malleret L, Fiorini F, Bossi L. 2009. Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes Dev* **23**: 2004–2015.
- Gudapaty S, Suzuki K, Wang X, Babitzke P, Romeo T. 2001. Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*. *J Bacteriol* **183**: 6017–6027.
- Guillier M, Gottesman S. 2006. Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. *Mol Microbiol* **59**: 231–247.
- Hollands K, Proshkin S, Sklyarova S, Epshtein V, Mironov A, Nudler E, Groisman EA. 2012. Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci* **109**: 5376–5381.
- Itoh Y, Rice JD, Goller C, Pannuri A, Taylor J, Meisner J, Beveridge TJ, Preston JF 3rd, Romeo T. 2008. Roles of *pgaABCD* genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-β-1,6-N-acetyl-D-glucosamine. *J Bacteriol* **190**: 3670–3680.
- Jackson DW, Suzuki K, Oakford L, Simecka JW, Hart ME, Romeo T. 2002. Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. *J Bacteriol* **184**: 290–301.
- Jorgensen MG, Thomason MK, Havelund J, Valentin-Hansen P, Storz G. 2013. Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev* **27**: 1132–1145.
- Lapouge K, Sineva E, Lindell M, Starke K, Baker CS, Babitzke P, Haas D. 2007. Mechanism of *hcnA* mRNA recognition in the Gac/Rsm signal transduction pathway of *Pseudomonas fluorescens*. *Mol Microbiol* **66**: 341–356.
- Lapouge K, Schubert M, Allain FH, Haas D. 2008. Gac/Rsm signal transduction pathway of γ-proteobacteria: from RNA recognition to regulation of social behaviour. *Mol Microbiol* **67**: 241–253.
- Lawhon SD, Frye JG, Suyemoto M, Porwollik S, McClelland M, Altier C. 2003. Global regulation by CsrA in *Salmonella typhimurium*. *Mol Microbiol* **48**: 1633–1645.
- Lemire S, Figueroa-Bossi N, Bossi L. 2011. Bacteriophage cross-talk: coordination of prophage induction by *trans*-acting antirepressors. *PLoS Genet* **7**: e1002149.
- Lenz DH, Miller MB, Zhu J, Kulkarni RV, Bassler BL. 2005. CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Mol Microbiol* **58**: 1186–1202.
- Liu MY, Romeo T. 1997. The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein. *J Bacteriol* **179**: 4639–4642.
- Macconkey A. 1905. Lactose-fermenting bacteria in faeces. *J Hyg (Lond)* **5**: 333–379.
- Martinez LC, Yakhnin H, Camacho MI, Georgellis D, Babitzke P, Puente JL, Bustamante VH. 2011. Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the *Salmonella* SPI-1 and SPI-2 virulence regulons through HilD. *Mol Microbiol* **80**: 1637–1656.

- Mercante J, Edwards AN, Dubey AK, Babitzke P, Romeo T. 2009. Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression. *J Mol Biol* **392**: 511–528.
- Miller JH. 1992. *A short course in bacterial genetics. a laboratory manual and handbook for Escherichia coli and related bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Mooney RA, Schweimer K, Rosch P, Gottesman M, Landick R. 2009. Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J Mol Biol* **391**: 341–358.
- Murphy KC, Campellone KG, Poteete AR. 2000. PCR-mediated gene replacement in *Escherichia coli*. *Gene* **246**: 321–330.
- Pan T, Sosnick T. 2006. RNA folding during transcription. *Annu Rev Biophys Biomol Struct* **35**: 161–175.
- Pannuri A, Yakhnin H, Vakulskas CA, Edwards AN, Babitzke P, Romeo T. 2012. Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. *J Bacteriol* **194**: 79–89.
- Patterson-Fortin LM, Vakulskas CA, Yakhnin H, Babitzke P, Romeo T. 2013. Dual posttranscriptional regulation via a co-factor-responsive mRNA leader. *J Mol Biol* **425**: 3662–3677.
- Pessi G, Williams F, Hindle Z, Heurlier K, Holden MT, Camara M, Haas D, Williams P. 2001. The global posttranscriptional regulator RsmA modulates production of virulence determinants and N-acylhomoserine lactones in *Pseudomonas aeruginosa*. *J Bacteriol* **183**: 6676–6683.
- Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. 2012. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev* **26**: 2621–2633.
- Plumbridge J, Bossi L, Oberto J, Wade JT, Figueroa-Bossi N. 2014. Interplay of transcriptional and small RNA-dependent control mechanisms regulates chitosugar uptake in *Escherichia coli* and *Salmonella*. *Mol Microbiol* doi: 10.1111/mmi.12573.
- Rabhi M, Espeli O, Schwartz A, Cayrol B, Rahmouni AR, Arluison V, Boudvillain M. 2011. The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J* **30**: 2805–2816.
- Romeo T, Gong M, Liu MY, Brun-Zinkernagel AM. 1993. Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol* **175**: 4744–4755.
- Romeo T, Vakulskas CA, Babitzke P. 2013. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ Microbiol* **15**: 313–324.
- Savageau MA. 1977. Design of molecular control mechanisms and the demand for gene expression. *Proc Natl Acad Sci* **74**: 5647–5651.
- Savageau MA. 1989. Are there rules governing patterns of gene regulation? In *Theoretical biology—epigenetic and evolutionary order from complex systems* [ed. Goodwin BC, Saunders PT], pp. 42–66. Edinburgh University Press, Edinburgh.
- Schmieger H. 1972. Phage P22-mutants with increased or decreased transduction abilities. *Mol Gen Genet* **119**: 75–88.
- Schubert M, Lapouge K, Duss O, Oberstrass FC, Jelezarov I, Haas D, Allain FH. 2007. Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat Struct Mol Biol* **14**: 807–813.
- Steiner T, Kaiser JT, Marinkovic S, Huber R, Wahl MC. 2002. Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J* **21**: 4641–4653.
- Sterzenbach T, Nguyen KT, Nuccio SP, Winter MG, Vakulskas CA, Clegg S, Romeo T, Baumbler AJ. 2013. A novel CsrA titration mechanism regulates fimbrial gene expression in *Salmonella typhimurium*. *EMBO J* **32**: 2872–2883.
- Suzuki K, Wang X, Weilbacher T, Pernestig AK, Melefors O, Georgellis D, Babitzke P, Romeo T. 2002. Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of *Escherichia coli*. *J Bacteriol* **184**: 5130–5140.
- Takeuchi K, Kiefer P, Reimann C, Keel C, Dubuis C, Rolli J, Vorholt JA, Haas D. 2009. Small RNA-dependent expression of secondary metabolism is controlled by Krebs cycle function in *Pseudomonas fluorescens*. *J Biol Chem* **284**: 34976–34985.
- Thomason MK, Fontaine F, De Lay N, Storz G. 2012. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* **84**: 17–35.
- Timmermans J, Van Melder L. 2009. Conditional essentiality of the *csrA* gene in *Escherichia coli*. *J Bacteriol* **191**: 1722–1724.
- Wang X, Dubey AK, Suzuki K, Baker CS, Babitzke P, Romeo T. 2005. CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Mol Microbiol* **56**: 1648–1663.
- Yakhnin H, Baker CS, Berezin I, Evangelista MA, Rassin A, Romeo T, Babitzke P. 2011. CsrA represses translation of *sdiA*, which encodes the N-acylhomoserine-L-lactone receptor of *Escherichia coli*, by binding exclusively within the coding region of *sdiA* mRNA. *J Bacteriol* **193**: 6162–6170.
- Yakhnin AV, Baker CS, Vakulskas CA, Yakhnin H, Berezin I, Romeo T, Babitzke P. 2013. CsrA activates *flhDC* expression by protecting *flhDC* mRNA from RNase E-mediated cleavage. *Mol Microbiol* **87**: 851–866.
- Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci* **97**: 5978–5983.
- Zou LL, Richardson JP. 1991. Enhancement of transcription termination factor *rho* activity with potassium glutamate. *J Biol Chem* **266**: 10201–10209.
- Zuker M. 2003. Mfold Web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

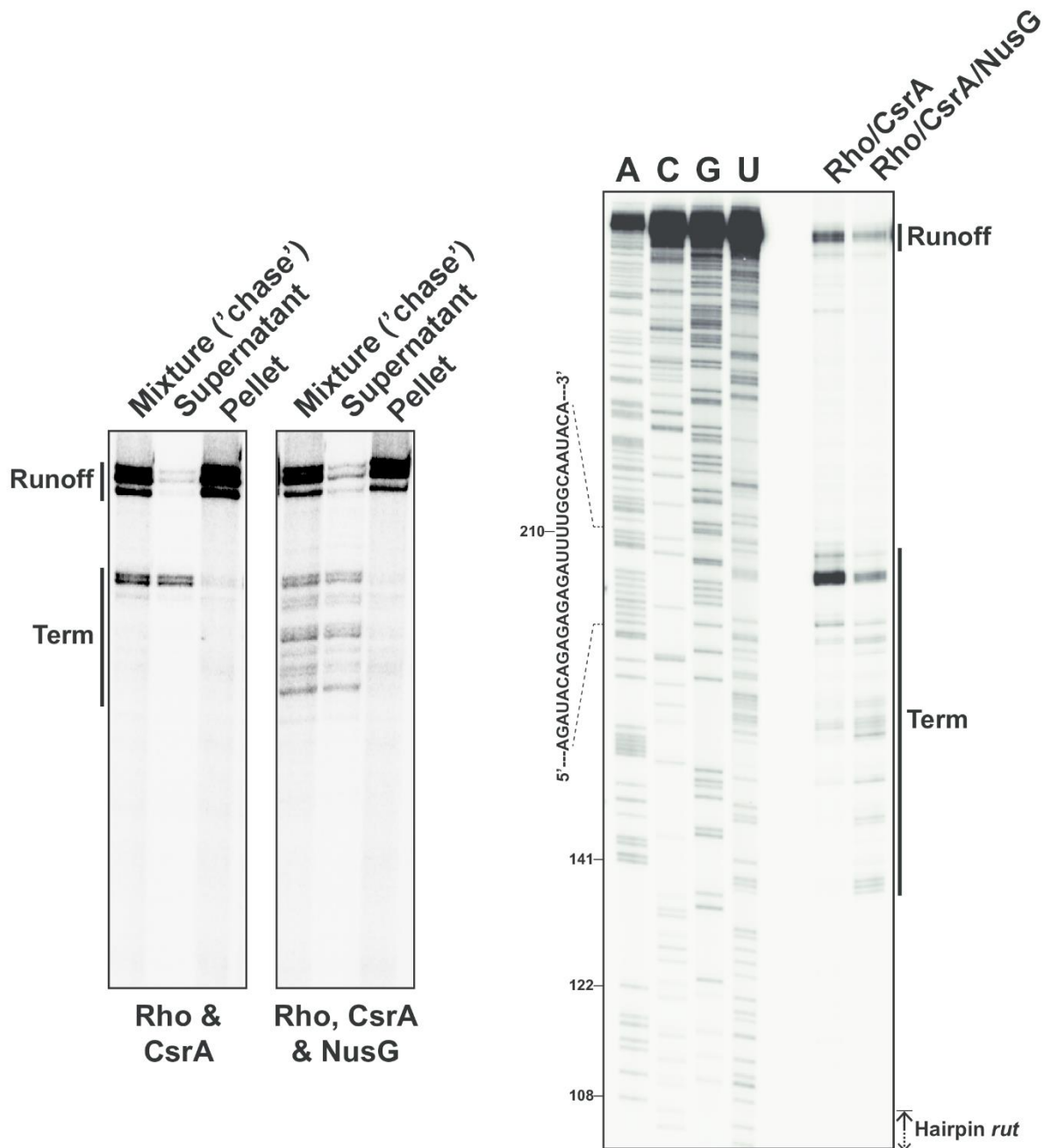
Supplementary Figures



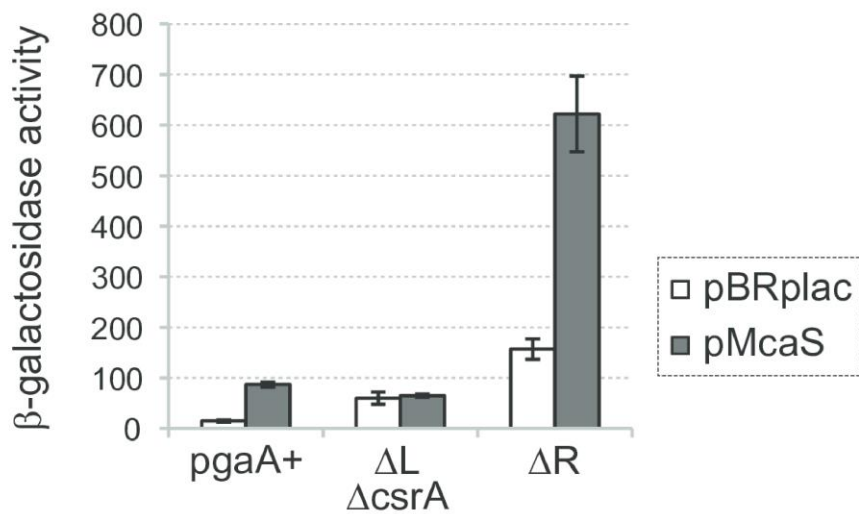
Supplemental Figure S1. Scale diagram of the *pgaA* gene and transcripts formed *in vivo* in absence (-Bcm, in blue) or presence (+Bcm, in red) of the Rho-inhibiting drug bicyclomycin (Peters et al. 2012). The shortest and longest *pgaA* transcripts released *in vitro* upon the action of Rho (termination window, in black) are also depicted for comparison (see text and Supplemental Figure S3 for details). Black lines delineate Bicyclomycin Significant Transcript (BST) regions as defined in Peters et al. 2012. Note that the *pgaABCD* operon is located on the minus strand of the *E. coli* genome.



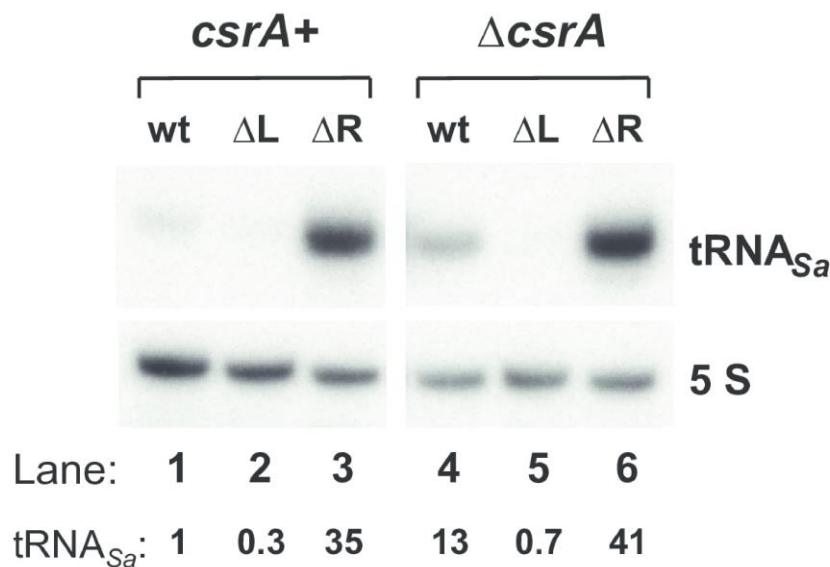
Supplemental Figure S2. Secondary structure of the *pgaA* leader mRNA as predicted by the *Mfold* algorithm (Zuker 2003). Components of the putative *rut* site and CsrA binding sites are highlighted in yellow and green, respectively. The presumptive McaS target sites are indicated by grey lines. The positions of the CsrA binding sites (core motifs) and McaS target sites are from (Wang et al. 2005; Thomason et al. 2012), respectively. We retrieved this *pgaA* leader sequence in more than fifty *E. coli* genomes (100% identity) from GenBank, suggesting that CsrA control of the Rho-dependent *pgaA* terminator is well conserved in this species. Intriguingly, we also found the *pgaA* leader (and *pgaA* gene) in the genome of *Shigella flexneri* strain 5 str. 8401 but not in other *Shigella* genomes. However, the sequence of the hairpin *rut* motif (positions 61-107)- or a structure with an equivalent role - could not be found upstream of *pgaA* genes in other Proteobacteria.



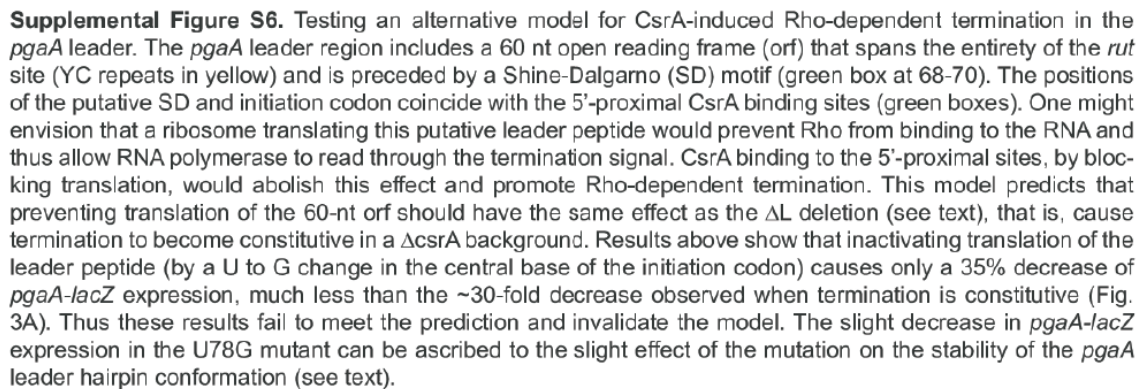
Supplemental Figure S3: Rho triggers *in vitro* transcription termination within the 5'-UTR of *pgaABCD* in the presence of CsrA and NusG. (Left) Single-round transcriptions with bead-affixed transcription complexes containing biotinylated 5'UTR-*pgaA* templates (Fig. 2A) (see Supplemental Methods for details). The 32P-labeled complexes, halted at +17 (by privation of CTP from the initiation mixture), were immobilized and purified on streptavidin-coated, magnetic beads (Ademabeads-300, Ademtech) before incubation with rNTPs and protein cofactors (indicated below gels). The supernatant and bead (pellet) fractions were then separated on a Magrack (GE-healthcare). The presence of the Rho-dependent transcripts in supernatants confirms that they stem from termination rather than transcriptional pausing or arrest. (Right) Identification of the main termination sites using phosphorothioate sequencing. Experiments were performed as described (Bossi et al. 2012). Note that in presence of Rho and CsrA, NusG shifts the start of the termination window towards promoter-proximal sites under both "chase" (left) and "standard" (right) transcription conditions.



Supplemental Figure S4: Overexpressing the *mcaS* gene. McaS activates *pgaA-lacZ* expression in the ΔR variant even though the latter lacks the presumptive McaS target site. Thus, the presumptive McaS target site is not needed for activation. Failure of McaS to activate *pgaA-lacZ* expression in the $\Delta L \Delta csrA$ strain is consistent with the finding that McaS acts by targeting CsrA (Jorgensen et al. 2013).



Supplemental Figure S5: Northern blot quantification of tRNA_{Sa} in strains expressing a constitutive (*CsrA*-irresponsive) *nhaR* gene. The $\Delta csrA$ deletion causes over a 10-fold increase in tRNA_{Sa} transcription in the *pgaA*⁺ strain (lanes 1 and 4) but has no significant effect in the ΔR leader variant (lanes 3 and 6). This shows that relieving *CsrA*-mediated repression of *NhaR* abolishes all of the effects of $\Delta csrA$ on the *pgaA* promoter (see text). tRNA_{Sa} counts were quantified by phosphorimaging, normalized to the 5S RNA and the values obtained divided by the value of lane 1.

[illegible]

Page 169

Supplemental Methods

Single-round “chase” transcription experiments

To monitor RNA release at the *pgaA* terminator, single-round transcription experiments were performed with bead-affixed transcription elongation complexes (Artsimovitch and Henkin 2009). DNA templates were prepared by PCR amplification of genomic DNA with forward primer *pgaA1* and a 5'-biotinylated variant of reverse primer *pgaA2* (Table S2). To allow preparation of transcription complexes halted at the +17 position of the leader sequence (by deprivation of CTP from the transcription initiation mixture), a C→T mutation at the +4 position was also introduced in the templates by PCR with mutagenic oligonucleotide *pgaA3* (Table S2). The biotinylated DNA templates (~10 nM, final concentration) were mixed with 10 μ L of streptavidin-coated magnetic beads (BioAdembeads-300, Ademtech) that had been pre-washed with activation buffer (1M KCl, 5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA) and incubated for 1h at room temperature under gentle stirring. Supernatants were then discarded and the beads washed thrice with modified transcription buffer (40 mM Tris-HCl, pH 7.9, 50 mM KCl, 2 mM MgCl₂, 0.1 mM DTT, 0.1 mg/mL BSA; 0.04 U/ μ L SUPERase-In). Then, beads were mixed with ApG (100 μ M), ATP and GTP (5 μ M each), UTP (0.8 μ M), 0.6 μ Ci/ μ L of [α -³²P]-UTP, and RNA polymerase (20 nM) in modified transcription buffer. After incubation for 20 min at 30°C, supernatants were eliminated and the beads were washed twice with “chase” transcription buffer (40 mM Tris-HCl, pH 7.9, 150 mM KCl, 5 mM MgCl₂, 0.1 mM DTT, 0.1 mg/mL BSA; 0.4 U/ μ L SUPERase-In). Then, beads were mixed with Rho (70 nM), CsrA (0 or 70 nM) and NusG (0 or 140 nM) in “chase” transcription buffer and incubated at 30°C for 10 min before addition of rNTPs (75 μ M, final concentrations) and rifampicin (25 μ g/mL, final concentration). After incubation for 5 min at 30°C, reactions were stopped by addition of 30 mM EDTA. Beads and supernatant fractions were separated, mixed with denaturing buffer (95% formamide, 30 mM EDTA), and analyzed by denaturing 8% polyacrylamide gel electrophoresis.

Supplemental References

- Artsimovitch I, Henkin TM. 2009. In vitro approaches to analysis of transcription termination. *Methods* **47**: 37-43.
- Bossi L, Schwartz A, Guillemardet B, Boudvillain M, Figueroa-Bossi N. 2012. A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes Dev* **26**: 1864-1873.
- Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640-6645.
- Figueroa-Bossi N, Coissac E, Netter P, Bossi L. 1997. Unsuspected prophage-like elements in *Salmonella typhimurium*. *Mol Microbiol* **25**: 161-173.
- Figueroa-Bossi N, Valentini M, Malleret L, Fiorini F, Bossi L. 2009. Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes Dev* **23**: 2004-2015.
- Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. 2012. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev* **26**: 2621-2633.
- Thomason MK, Fontaine F, De Lay N, Storz G. 2012. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* **84**: 17-35.
- Wang X, Dubey AK, Suzuki K, Baker CS, Babitzke P, Romeo T. 2005. CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Mol Microbiol* **56**: 1648-1663.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.

2. DISCUSSION

Les travaux effectués durant cette thèse nous ont permis de mieux évaluer l'importance de la terminaison Rho-dépendante au sein de la diversité bactérienne. Cette voie de régulation de la transcription a été étudiée essentiellement chez *E. coli* où Rho est impliqué dans plusieurs processus clés du métabolisme de l'ARN. Cette implication fait de Rho une protéine essentielle pour le développement d'*E. coli* (et des espèces apparentées) et une cible thérapeutique potentiellement intéressante.

Il est communément admis que le caractère essentiel de Rho n'est pas général dans le règne bactérien, notamment chez les espèces Gram⁺. Cette règle doit cependant être considérée avec précaution, du moins pour les Actinobactéries. Comme indiqué précédemment, le statut de protéine accessoire attribué au facteur Rho chez ce phylum découle principalement de tests de résistance à la BCM (voir page 97). Or ces derniers ont été réalisés sur trop peu d'espèces – en grande partie issues de l'ordre des *Corynebacteriales* (Miyoshi *et al.*, 1972; Nishida *et al.*, 1972) – pour être représentatifs, et le seul contre-exemple de bactérie Gram⁺ sensible à l'ATB se trouve être une Actinobactérie (*Micrococcus luteus*) (Nishida *et al.*, 1972; Nowatzke *et al.*, 1997b). Notre étude phylogénique présentée dans l'Article I nous a permis d'observer que *rho* est présent dans tous les génomes d'Actinobactéries, ce qui indique que la résistance à la BCM n'est pas due à une perte du gène au cours de l'évolution comme cela a sûrement été le cas pour le phylum des Ténérécutes (discutée ci-dessous). Par ailleurs, il est intéressant de noter que certaines Actinobactéries, notamment du genre *Streptomyces* (dont certains représentants sont les producteurs naturels de la BCM), possèdent une duplication du gène *rho* qui semble coder une protéine possédant des mutations lui conférant une résistance à la BCM. L'expression de ce paralogue pourrait garantir la présence d'un facteur Rho fonctionnel pour le développement de la bactérie, même en présence d'ATB.

Une autre hypothèse expliquant la résistance à la BCM serait que, chez certaines espèces, la terminaison Rho-dépendante utilise un mécanisme alternatif insensible aux effets de l'ATB. Ceci semblait être le cas du facteur Rho de *M. tuberculosis* qui était décrit dans une étude récente comme étant capable de provoquer la dissociation de CET artificiellement arrêtés alors que l'ATP avait été remplacé par des analogues non hydrolysables (Kalarickal *et al.*, 2010). Notre étude publiée dans l'Article II contredit la présence d'un tel mécanisme et, au contraire, supporte l'idée que MtbRho utilise un mécanisme similaire à celui d'EcRho,

c'est à dire que la dissociation du CET est dépendante de la capacité de l'enzyme à se fixer sur l'ARN et à se transloquer dans la direction 5'→3' de manière ATP-dépendante. De plus, nous avons observé que cette activité de MtbRho peut être inhibée par la BCM ($IC_{50MtbRho} \sim 280 \mu M$), même s'il est vrai que les doses inhibitrices sont très supérieures à celles requises pour EcRho ($IC_{50EcRho} \sim 30 \mu M$). Malheureusement, nous n'avons pas pu déterminer l'origine exacte de cette différence de sensibilité à l'ATB et plusieurs hypothèses restent possibles (voir ci-dessous). Néanmoins, s'il s'avérait que la raison sous-jacente est répandue chez les Actinobactéries, alors cela expliquerait pourquoi la plupart des espèces testées montrent une résistance à la BCM *in vivo*. Cela impliquerait que les Actinobactéries ont développé divers mécanismes de résistance afin de protéger la terminaison Rho-dépendante des effets de la BCM (duplication du gène, évolution du facteur) ce qui indiquerait que le facteur Rho est une enzyme importante dans ce phylum contrairement à ce qui a été décrit jusqu'à maintenant. L'implication très récente du facteur Rho dans divers mécanismes de régulation du métabolisme de *Corynebacterium glutamicum* (Takemoto *et al.*, 2015; Tanaka *et al.*, 2015) conforte cette hypothèse. Cette idée est également supportée par le fait que, lors de cribles génomiques, *rho* n'est pas retrouvé dans le pool des gènes accessoires dont l'inactivation n'empêche pas la croissance de *M. tuberculosis* (Griffin *et al.*, 2011; Sassetti *et al.*, 2003).

Au-delà de sa faible sensibilité à la BCM, MtbRho possède d'autres caractéristiques qui nous paraissent intrigantes. Tout comme les études précédentes (Kalarickal *et al.*, 2010; Mitra *et al.*, 2014), nous avons observé que MtbRho a une (relativement) faible activité NTPase et une préférence prononcée pour les purines-triphosphates. D'autre part, nous avons démontré que MtbRho possède bien une activité hélicase mais uniquement avec de courts duplexes (14 pb). Chez *E. coli*, l'activité hélicase de Rho est vraisemblablement cruciale pour dissocier (résoudre) les R-Loops. La formation de telles structures chez *M. tuberculosis* n'a pas encore été établie avec certitude même si elle est supposée être plus importante pour les génomes riches en G/C (Roy & Lieber, 2009). Nous supposons que d'autres facteurs (couplage traduction/transcription plus efficace, activité RNase H ou topoisomérase plus marquée, etc.) compensent l'activité hélicase modeste de MtbRho si la formation de R-Loops transcriptionnelles doit également être contrée chez *M. tuberculosis*. Ce dernier point est supporté par l'observation récente que la RNase H est une enzyme essentielle chez l'espèce apparentée *Mycobacterium smegmatis* (Minias *et al.*, 2015).

Nous avons identifié deux raisons possibles pour expliquer l'activité hélicase modeste de MtbRho. La première est un amoindrissement du réseau catalytique d'interactions ARN-

SBS dû notamment à la substitution d'un résidu basique clé de la R-Loop (K326 chez EcRho) par une thréonine (T501 chez MtbRho). Il est intéressant de noter que cette substitution est probablement une synapomorphie (*"caractère dérivé et partagé par l'ensemble des taxons d'un groupe monophylétique et qui a donc été acquis par leur ancêtre commun"*) pour l'ordre des *Corynebacteriales* (**Figure 45**).

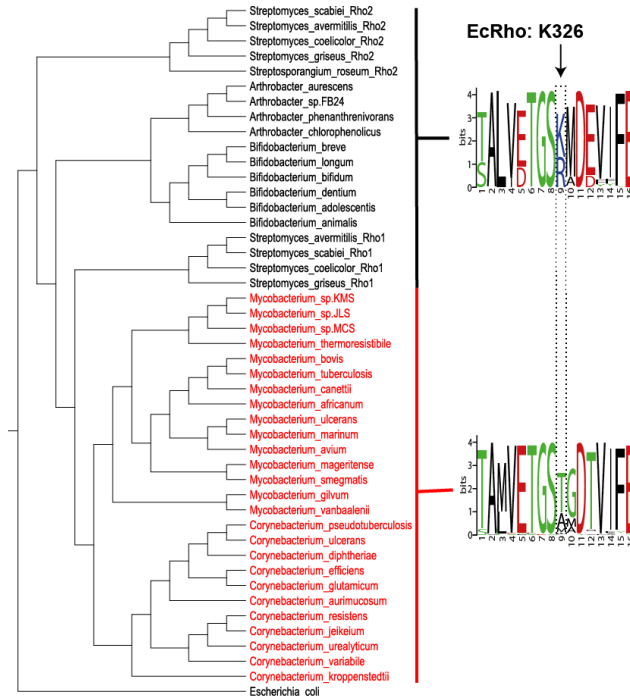


Figure 45 : Conservation de la R-Loop chez les Actinobactéries.
La lysine habituellement retrouvée dans la R-Loop des facteurs Rho est substituée par une thréonine chez les *Corynebacteriales* (rouge).

Cette substitution Lys→Thr semble également marquer une inversion dans l'ordre d'importance fonctionnelle des composants du SBS : R-Loop > Q-Loop pour EcRho (et, sans doute, pour beaucoup d'espèces) (Ma & Schulten, 2015) mais Q-Loop > R-Loop pour MtbRho (et, possiblement, pour les autres *Corynebacteriales*). Cette inversion, supportée par les données existantes de mutagenèse (voir Article II), pourrait contribuer à expliquer la faible sensibilité de MtbRho à la BCM. Cette proposition découle du fait que, si la BCM inhibe l'hydrolyse de l'ATP, elle semble également affecter l'interaction avec l'ARN, probablement parce qu'elle-même interagit avec la R-Loop (voir page 98) (Magyar *et al.*, 1996; Skordalakes *et al.*, 2005). Les facteurs Rho pour lesquels la Q-Loop (et non la R-Loop) contribue majoritairement au réseau d'interactions SBS:ARN pourraient donc s'avérer moins sensibles à la BCM. Il est possible également que la substitution Lys→Thr altère le positionnement de la R-Loop et son interaction avec la BCM (voir page 98), de ce fait diminuant l'affinité de Rho pour l'ATB.

La deuxième raison pouvant expliquer l'activité hélicase modeste de MtbRho est de nature stérique et liée à la présence de l'insertion de 143 acides aminés dans le NTD de

MtbRho. La délétion de ce domaine d'insertion s'accompagne d'une augmentation de la processivité de l'hélicase suggérant que la présence de cet appendice structural dans chaque sous-unité (voir la **Figure 1** de l'Article II, page 134) favorise l'ouverture de l'anneau hexamérique (et, donc, le relargage du substrat ARN). Néanmoins, un effet négatif de cette délétion est qu'elle diminue l'affinité de MtbRho pour l'ARN et sa capacité à provoquer une terminaison de la transcription précoce. Cette dernière observation est en accord avec l'idée que les insertions NTD retrouvées dans les facteurs Rho d'Actinobactéries facilitent l'ancrage de ces enzymes aux sites *Rut* de transcrits inévitablement plus structurés du fait des taux G/C élevés des génomes (et, donc, des transcriptomes) de ces espèces (Ingham *et al.*, 1996; Nowatzke *et al.*, 1997a). En d'autres termes, ils jouent un peu le rôle du facteur NusG qui facilite l'utilisation de terminateurs imparfaits par EcRho (voir page 84). On peut d'ailleurs noter que MtbRho est complètement insensible à l'action de NusG (Kalarickal *et al.*, 2010), ce qui ne semble pas être le cas du mutant de délétion Δ MtbRho (A. Schwartz & M. Boudvillain, données préliminaires non publiées).

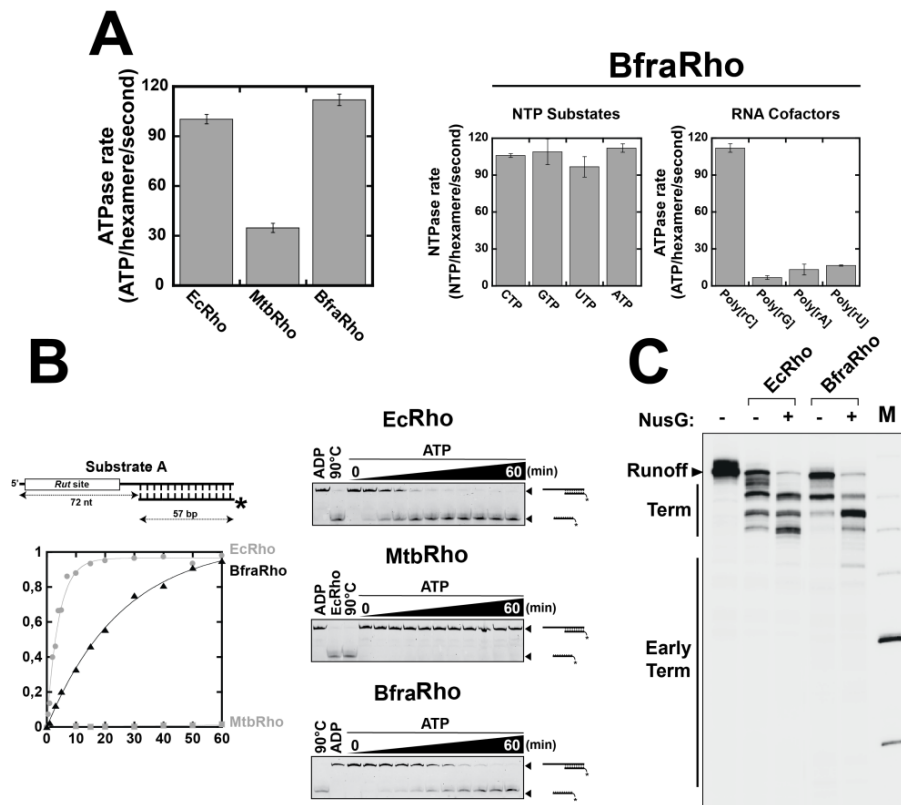


Figure 46 : Activités enzymatiques du facteur Rho de *Bacteroides fragilis*. (A) BfraRho et EcRho ont des activités NTPase similaires, peu affectées par la nature du NTP. (B) Comme EcRho (mais pas MtbRho), BfraRho peut dissocier un duplexe ARN:ADN de 57 pb. (C) BfraRho est capable d'utiliser le terminateur Rho-dépendant λ rI d'une manière similaire à EcRho mais ne provoque pas de terminaison précoce (Early Term). Les conditions expérimentales sont celles de l'Article III.

Notre étude phylogénique (cf. Article I) a révélé que les Actinobactéries n'étaient pas le seul phylum à posséder des insertions NTD. On retrouve notamment de telles insertions dans de nombreux facteurs Rho de Bactéroidetes bien que leurs compositions soient différentes (plutôt riches en Gln/Asn chez les Bactéroidetes et en Arg/Lys chez les Actinobactéries). Il est possible que cette différence de composition soit liée au fait que les génomes des Bactéroidetes sont comparativement riches en A/T. Dans une étude préliminaire des activités enzymatiques du facteur Rho de *Bacteroides fragilis* (BfraRho), nous avons observé que ce facteur se comporte de manière similaire à EcRho et n'avons pas pu mettre en évidence de rôle notable pour le domaine d'insertion NTD (**Figure 46**). Notamment, BfraRho n'induit pas de terminaison précoce et est sensible à l'action du facteur NusG d'*E. coli* (**Figure 46**).

Un aspect notable de nos travaux est qu'ils confortent l'idée que la terminaison Rho-dépendante est un élément important des mécanismes de régulation conditionnelle de l'expression génique. Premièrement, nous ne retrouvons pas de gène *rho* dans les génomes d'espèces appartenant au phylum des Ténéricutes, un embranchement composé quasi exclusivement de bactéries parasites ou symbiontes. Ces modes de vie s'accompagnent souvent d'une évolution réductive du génome caractérisée par une perte de gènes devenus obsolètes, notamment ceux impliqués dans la régulation génique (Andersson & Kurland, 1998; Fraser *et al.*, 1995). Plus généralement, nous avons observé une corrélation positive entre la taille des génomes ou leur taux de C/G (deux paramètres interdépendants généralement associés au degré de complexité de l'écosystème) et la présence de *rho*. Ainsi, *rho* est absent chez des espèces bactériennes occupant des niches écologiques restreintes (présentant peu de variations, comme pour les Ténéricutes cités ci-dessus) et qui donc n'imposent pas une haute régulation de l'expression génique. Ce type de caractéristiques (génome réduit, faible taux de G/C) est également supposé rendre l'action du régulateur Hfq superflue chez certaines espèces appartenant notamment aux Cyanobactéries et aux Firmicutes (Jousselin *et al.*, 2009), deux phyla pour lesquels, respectivement, tous ou une fraction des génomes analysés sont dépourvus du gène *rho*. Cette ligne de raisonnement soulève néanmoins deux questions : *i*) Pourquoi certaines espèces intracellulaires, notamment de l'ordre des *Rickettsiales* (α -Protéobactéries) possèdent une, voire deux copies du gène *rho* ? *ii*) Comment les espèces dépourvues de facteur Rho gèrent-elles la transcription antisense, la résolution des R-Loops et la répression des gènes de bactériophages, trois

fonctions cruciales et impliquant Rho chez *E. coli* (Cardinale *et al.*, 2008; Leela *et al.*, 2013; Peters *et al.*, 2012).

Deuxièmement, nous avons démontré que Rho est impliqué dans la régulation de l'opéron *pgaABCD* d'*E. coli* (Article III). L'expression de *pgaABCD* est importante pour la formation de biofilm (Wang *et al.*, 2004), un état facilitant le développement bactérien dans des environnements représentant un haut niveau de stress. Nos travaux démontrent que l'expression de cet opéron est réprimée par un mécanisme impliquant le facteur Rho ainsi que le régulateur CsrA. Jusqu'ici, CsrA était décrit pour agir essentiellement au niveau post-transcriptionnel en entrant en compétition avec le ribosome pour la fixation sur le site *rbs*. Cependant, notre étude démontre que CsrA peut aussi avoir une action au niveau de la transcription. En se fixant dans la partie 5'-leader de l'ARNm *pgaA*, cette RBP provoque le remodelage de l'ARN et permet à Rho d'accéder à un site *Rut* qui était emprisonné dans une structure tige-boucle. Cette découverte conforte l'idée que chez *E. coli* (et éventuellement chez les espèces phylo-apparentées) la terminaison Rho-dépendante est associée à de nombreux mécanismes de régulation de l'expression génique que ce soit via l'action de Hfq (Rabhi *et al.*, 2011a), de sRNA (Bossi *et al.*, 2012; Wang *et al.*, 2015), de H-NS (Kotlajich *et al.*, 2015; Peters *et al.*, 2012), ou de motifs riboswitch (Hollands *et al.*, 2012; Takemoto *et al.*, 2015). Nous pronostiquons que de nouveaux mécanismes régulateurs Rho-dépendants impliquant d'autres RBP, PyrR par exemple (Tanaka *et al.*, 2015), seront rapidement mis en évidence.

3. CONCLUSION/PERSPECTIVES

Ces travaux de thèse nous ont permis de mieux évaluer la distribution du facteur Rho au sein de la biodiversité. Rho apparaît être une enzyme présente chez la grande majorité des phyla bactériens, ce qui sous-entend une relative importance de la terminaison Rho-dépendante au sein de la diversité bactérienne. Cette importance semble être liée à la capacité de la bactérie à réguler l'expression de son génome, dans l'optique de s'adapter aux différentes variations de son écosystème. Cette relation est bien illustrée par notre étude de la régulation de l'opéron *pgaABCD* d'*E. coli*, impliqué dans la formation de biofilm. Puisque la terminaison Rho-dépendante semble être un mécanisme important pour l'adaptation rapide de la bactérie à son environnement, il serait intéressant d'étudier l'influence de Rho dans différentes conditions de culture ou de stress pour la bactérie. Ce genre d'études pourrait

s'avérer particulièrement pertinent pour des espèces chez qui il est actuellement supposé que Rho n'est pas essentiel.

Nos travaux sur le facteur Rho de *M. tuberculosis* suggèrent que le mécanisme général de fonctionnement de cette enzyme (dont le prototype est EcRho) est conservé au travers la biodiversité. Néanmoins, certaines caractéristiques de MtbRho restent encore obscures, notamment le fait que son inhibition nécessite de hautes doses de BCM. Déterminer les raisons de cette faible sensibilité à l'ATB pourrait aider à mieux comprendre le spectre d'action de la BCM, notamment chez les Actinobactéries. Dans ce contexte, il pourrait être particulièrement intéressant de mener des études structurales sur le facteur MtbRho, même si les essais de cristallisation menés en collaboration avec l'équipe de Bertrand Castaing se sont pour l'instant révélés peu concluants. La structure de MtbRho et sa comparaison avec celles obtenues pour EcRho (Canals *et al.*, 2010; Skordalakes & Berger, 2003; Skordalakes *et al.*, 2005; Skordalakes & Berger, 2006; Thomsen & Berger, 2009) pourraient également permettre de mieux comprendre les différences de comportement enzymatique observées (cf. Article II) et de déterminer le rôle du domaine d'insertion NTD de MtbRho. Enfin, si nous avons maintenant une meilleure idée des mécanismes utilisés par MtbRho *in vitro*, nous n'avons encore que très peu d'informations sur son fonctionnement dans un contexte physiologique. Nous ne connaissons pas la composition de ses sites *Rut*, l'identité (et la part) des gènes/opérons qu'elle régule, la nature de ses interactions avec les composants et cofacteurs de la machinerie de transcription, ou encore si MtbRho est impliqué dans la résolution des R-Loops, l'inhibition de la transcription antisense, la répression des xénogènes, et/ou d'autres fonctions propres à *M. tuberculosis*. Évaluer l'importance de la terminaison Rho-dépendante chez cette espèce pathogène pourrait s'avérer important dans l'optique d'utiliser Rho comme cible thérapeutique. Toujours dans cette optique, les divers essais enzymatiques présentés dans ce mémoire devraient constituer des bases de départ intéressantes pour le développement de crible(s) haut-débit de chimiothèques à la recherche de nouveaux inhibiteurs du facteur Rho.

Bibliographie

- Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. (2005).** Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460-465.
- Abe, H., Abo, T. & Aiba, H. (1999).** Regulation of intrinsic terminator by translation in *Escherichia coli*: transcription termination at a distance downstream. *Genes Cells* **4**, 87-97.
- Adelman, J. L., Jeong, Y. J., Liao, J. C., Patel, G., Kim, D. E., Oster, G. & Patel, S. S. (2006).** Mechanochemistry of transcription termination factor Rho. *Mol Cell* **22**, 611-621.
- Adhya, S. & Gottesman, M. (1978).** Control of transcription termination. *Annu Rev Biochem* **47**, 967-996.
- Agrawal, A., Mohanty, B. K. & Kushner, S. R. (2014).** Processing of the seven valine tRNAs in *Escherichia coli* involves novel features of RNase P. *Nucleic Acids Res* **42**, 11166-11179.
- Alifano, P., Rivellini, F., Limauro, D., Bruni, C. B. & Carlomagno, M. S. (1991).** A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* **64**, 553-563.
- Altman, S. (1971).** Isolation of tyrosine tRNA precursor molecules. *Nat New Biol* **229**, 19-21.
- Altman, S. & Smith, J. D. (1971).** Tyrosine tRNA precursor molecule polynucleotide sequence. *Nat New Biol* **233**, 35-39.
- Anantharaman, V., Koonin, E. V. & Aravind, L. (2002).** Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res* **30**, 1427-1464.
- Andersson, S. G. & Kurland, C. G. (1998).** Reductive evolution of resident genomes. *Trends Microbiol* **6**, 263-268.
- Appleby, T. C., Anderson, R., Fedorova, O., Pyle, A. M., Wang, R., Liu, X., Brendza, K. M. & Somoza, J. R. (2011).** Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. *J Mol Biol* **405**, 1139-1153.
- Arnvig, K. B., Zeng, S., Quan, S., Papageorge, A., Zhang, N., Villapakkam, A. C. & Squires, C. L. (2008).** Evolutionary comparison of ribosomal operon antitermination function. *J Bacteriol* **190**, 7251-7257.
- Artsimovitch, I. & Landick, R. (1998).** Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. *Genes Dev* **12**, 3110-3122.
- Artsimovitch, I. & Landick, R. (2000).** Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**, 7090-7095.
- Awano, N., Xu, C., Ke, H., Inoue, K., Inouye, M. & Phadtare, S. (2007).** Complementation analysis of the cold-sensitive phenotype of the *Escherichia coli* *csdA* deletion strain. *J Bacteriol* **189**, 5808-5815.
- Baker, C. S., Morozov, I., Suzuki, K., Romeo, T. & Babitzke, P. (2002).** CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. *Mol Microbiol* **44**, 1599-1610.
- Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. (2012).** Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol Microbiol* **85**, 21-38.
- Bakshi, S., Choi, H. & Weisshaar, J. C. (2015).** The spatial biology of transcription and translation in rapidly growing *Escherichia coli*. *Front Microbiol* **6**, 636.
- Balasubramanian, K. & Stitt, B. L. (2010).** Evidence for amino acid roles in the chemistry of ATP hydrolysis in *Escherichia coli* Rho. *J Mol Biol* **404**, 587-599.
- Bandyra, K. J., Bouvier, M., Carpousis, A. J. & Luisi, B. F. (2013).** The social fabric of the RNA degradosome. *Biochim Biophys Acta* **1829**, 514-522.
- Banroques, J., Cordin, O., Doère, M., Linder, P. & Tanner, N. K. (2008).** A conserved phenylalanine of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins. *Mol Cell Biol* **28**, 3359-3371.
- Banroques, J., Doère, M., Dreyfus, M., Linder, P. & Tanner, N. K. (2010).** Motif III in superfamily 2 "helicases" helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J Mol Biol* **396**, 949-966.
- Barik, S., Bhattacharya, P. & Das, A. (1985).** Autogenous regulation of transcription termination factor Rho. *J Mol Biol* **182**, 495-508.
- Barria, C., Malecki, M. & Arraiano, C. M. (2013).** Bacterial adaptation to cold. *Microbiology* **159**, 2437-2443.
- Barrick, J. E. & Breaker, R. R. (2007).** The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8**, R239.
- Bear, D. G., Andrews, C. L., Singer, J. D., Morgan, W. D., Grant, R. A., von Hippel, P. H. & Platt, T. (1985).** *Escherichia coli* transcription termination factor rho has a two-domain structure in its activated form. *Proc Natl Acad Sci U S A* **82**, 1911-1915.
- Belogurov, G. A., Vassilyeva, M. N., Svetlov, V., Klyuyev, S., Grishin, N. V., Vassilyev, D. G. & Artsimovitch, I. (2007).** Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell* **26**, 117-129.
- Bennett, C. L. & La Spada, A. R. (2015).** Unwinding the role of senataxin in neurodegeneration. *Discov Med* **19**, 127-136.
- Berger, J. M. (2008).** SnapShot: nucleic acid helicases and translocases. *Cell* **134**, 888-888 e881.

- Bernstein, J. A., Lin, P. H., Cohen, S. N. & Lin-Chao, S. (2004).** Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proc Natl Acad Sci U S A* **101**, 2758-2763.
- Bizebard, T., Ferlenghi, I., Iost, I. & Dreyfus, M. (2004).** Studies on three E. coli DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry* **43**, 7857-7866.
- Blombach, F., Daviter, T., Fielden, D., Grohmann, D., Smollett, K. & Werner, F. (2013).** Archaeology of RNA polymerase: factor swapping during the transcription cycle. *Biochem Soc Trans* **41**, 362-367.
- Blumenthal, R. M., Reeh, S. & Pedersen, S. (1976).** Regulation of transcription factor rho and the alpha subunit of RNA polymerase in Escherichia coli B/r. *Proc Natl Acad Sci U S A* **73**, 2285-2288.
- Bogden, C. E., Fass, D., Bergman, N., Nichols, M. D. & Berger, J. M. (1999).** The structural basis for terminator recognition by the Rho transcription termination factor. *Mol Cell* **3**, 487-493.
- Bono, F., Ebert, J., Lorentzen, E. & Conti, E. (2006).** The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713-725.
- Borukhov, S., Sagitov, V. & Goldfarb, A. (1993).** Transcript cleavage factors from E. coli. *Cell* **72**, 459-466.
- Borukhov, S., Lee, J. & Laptenko, O. (2005).** Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol* **55**, 1315-1324.
- Bossi, L., Schwartz, A., Guillemardet, B., Boudvillain, M. & Figueroa-Bossi, N. (2012).** A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes Dev* **26**, 1864-1873.
- Boudvillain, M., Nollmann, M. & Margeat, E. (2010).** Keeping up to speed with the transcription termination factor Rho motor. *Transcription* **1**, 70-75.
- Boudvillain, M., Figueroa-Bossi, N. & Bossi, L. (2013).** Terminator still moving forward: expanding roles for Rho factor. *Curr Opin Microbiol* **16**, 118-124.
- Brennan, C. A., Dombroski, A. J. & Platt, T. (1987).** Transcription termination factor rho is an RNA-DNA helicase. *Cell* **48**, 945-952.
- Brennan, C. A., Steinmetz, E. J., Spear, P. & Platt, T. (1990).** Specificity and efficiency of rho-factor helicase activity depends on magnesium concentration and energy coupling to NTP hydrolysis. *J Biol Chem* **265**, 5440-5447.
- Brenneis, M., Hering, O., Lange, C. & Soppa, J. (2007).** Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea. *PLoS Genet* **3**, e229.
- Briercheck, D. M., Wood, T. C., Allison, T. J., Richardson, J. P. & Rule, G. S. (1998).** The NMR structure of the RNA binding domain of E. coli rho factor suggests possible RNA-protein interactions. *Nat Struct Biol* **5**, 393-399.
- Britton, R. A. (2009).** Role of GTPases in bacterial ribosome assembly. *Annu Rev Microbiol* **63**, 155-176.
- Broeze, R. J., Solomon, C. J. & Pope, D. H. (1978).** Effects of low temperature on in vivo and in vitro protein synthesis in Escherichia coli and Pseudomonas fluorescens. *J Bacteriol* **134**, 861-874.
- Brosh, R. M. (2013).** DNA helicases involved in DNA repair and their roles in cancer. *Nat Rev Cancer* **13**, 542-558.
- Brown, S., Albrechtsen, B., Pedersen, S. & Klemm, P. (1982).** Localization and regulation of the structural gene for transcription-termination factor rho of Escherichia coli. *J Mol Biol* **162**, 283-298.
- Browne, R. J., Barr, E. W. & Stitt, B. L. (2005).** Catalytic cooperativity among subunits of Escherichia coli transcription termination factor Rho. Kinetics and substrate structural requirements. *J Biol Chem* **280**, 13292-13299.
- Browne, R. J. & Stitt, B. L. (2005).** Active site occupancy required for catalytic cooperativity by Escherichia coli transcription termination factor Rho. *J Biol Chem* **280**, 13300-13303.
- Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E. & Rösch, P. (2010).** A NusE:NusG complex links transcription and translation. *Science* **328**, 501-504.
- Burns, C. M., Richardson, L. V. & Richardson, J. P. (1998).** Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J Mol Biol* **278**, 307-316.
- Burns, C. M., Nowatzke, W. L. & Richardson, J. P. (1999).** Activation of Rho-dependent transcription termination by NusG. Dependence on terminator location and acceleration of RNA release. *J Biol Chem* **274**, 5245-5251.
- Byrd, A. K. & Raney, K. D. (2004).** Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. *Nat Struct Mol Biol* **11**, 531-538.
- Byrd, A. K. & Raney, K. D. (2012).** Superfamily 2 helicases. *Front Biosci (Landmark Ed)* **17**, 2070-2088.
- Caldelari, I., Chao, Y., Romby, P. & Vogel, J. (2013).** RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harb Perspect Med* **3**, a010298.
- Canals, A., Uson, I. & Coll, M. (2010).** The structure of RNA-free Rho termination factor indicates a dynamic mechanism of transcript capture. *J Mol Biol* **400**, 16-23.
- Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. & Nudler, E. (2008).** Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* **320**, 935-938.

- Carlomagno, M. S. & Nappo, A. (2003).** NusA modulates intragenic termination by different pathways. *Gene* **308**, 115-128.
- Cartier, G., Lorieux, F., Allemand, F., Dreyfus, M. & Bizebard, T. (2010).** Cold adaptation in DEAD-box proteins. *Biochemistry* **49**, 2636-2646.
- Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000).** Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A* **97**, 13080-13085.
- Celesnik, H., Deana, A. & Belasco, J. G. (2007).** Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal. *Mol Cell* **27**, 79-90.
- Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H. & Conti, E. (2011).** Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell* **41**, 693-703.
- Chalissery, J., Banerjee, S., Bandey, I. & Sen, R. (2007).** Transcription termination defective mutants of Rho: role of different functions of Rho in releasing RNA from the elongation complex. *J Mol Biol* **371**, 855-872.
- Chalissery, J., Muteeb, G., Kalarickal, N. C., Mohan, S., Jisha, V. & Sen, R. (2011).** Interaction surface of the transcription terminator Rho required to form a complex with the C-terminal domain of the antiterminator NusG. *J Mol Biol* **405**, 49-64.
- Chamot, D. & Owttrim, G. W. (2000).** Regulation of cold shock-induced RNA helicase gene expression in the Cyanobacterium anabaena sp. strain PCC 7120. *J Bacteriol* **182**, 1251-1256.
- Chan, C. L., Wang, D. & Landick, R. (1997).** Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *J Mol Biol* **268**, 54-68.
- Chandran, V., Poljak, L., Vanzo, N. F. & other authors (2007).** Recognition and cooperation between the ATP-dependent RNA helicase RhlB and ribonuclease RNase E. *J Mol Biol* **367**, 113-132.
- Chang, B., Halgamuge, S. & Tang, S. L. (2006).** Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. *Gene* **373**, 90-99.
- Charollais, J., Pflieger, D., Vinh, J., Dreyfus, M. & Iost, I. (2003).** The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in Escherichia coli. *Mol Microbiol* **48**, 1253-1265.
- Charollais, J., Dreyfus, M. & Iost, I. (2004).** CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Res* **32**, 2751-2759.
- Chen, X. & Stitt, B. L. (2009).** ADP but not P(i) dissociation contributes to rate limitation for Escherichia coli Rho. *J Biol Chem* **284**, 33773-33780.
- Chen, Y., Potratz, J. P., Tijerina, P., Del Campo, M., Lambowitz, A. M. & Russell, R. (2008).** DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc Natl Acad Sci U S A* **105**, 20203-20208.
- Chi, W., He, B., Manavski, N., Mao, J., Ji, D., Lu, C., Rochaix, J. D., Meurer, J. & Zhang, L. (2014).** RHON1 mediates a Rho-like activity for transcription termination in plastids of Arabidopsis thaliana. *Plant Cell* **26**, 4918-4932.
- Ciampi, M. S. (2006).** Rho-dependent terminators and transcription termination. *Microbiology* **152**, 2515-2528.
- Condon, C. & Putzer, H. (2002).** The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res* **30**, 5339-5346.
- Cordin, O., Tanner, N. K., Doère, M., Linder, P. & Banroques, J. (2004).** The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* **23**, 2478-2487.
- Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. (2006).** The DEAD-box protein family of RNA helicases. *Gene* **367**, 17-37.
- Czyz, A., Mooney, R. A., Iaconi, A. & Landick, R. (2014).** Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* **5**, e00931.
- D'Heygere, F., Rabhi, M. & Boudvillain, M. (2013).** Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology* **159**, 1423-1436.
- Dangkulwanich, M., Ishibashi, T., Bintu, L. & Bustamante, C. (2014).** Molecular mechanisms of transcription through single-molecule experiments. *Chem Rev* **114**, 3203-3223.
- Deana, A. & Belasco, J. G. (2005).** Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* **19**, 2526-2533.
- Deana, A., Celesnik, H. & Belasco, J. G. (2008).** The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**, 355-358.
- Delagoutte, E. & von Hippel, P. H. (2002).** Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I: Structures and properties of isolated helicases. *Q Rev Biophys* **35**, 431-478.
- Deneke, C., Lipowsky, R. & Valleriani, A. (2013).** Effect of ribosome shielding on mRNA stability. *Phys Biol* **10**, 046008.
- Dennis, P. P. & Nomura, M. (1974).** Stringent control of ribosomal protein gene expression in Escherichia coli. *Proc Natl Acad Sci U S A* **71**, 3819-3823.
- Dennis, P. P., Ehrenberg, M., Fange, D. & Bremer, H. (2009).** Varying rate of RNA chain elongation during rm transcription in Escherichia coli. *J Bacteriol* **191**, 3740-3746.

- Ding, S. C. & Pyle, A. M. (2012). Molecular mechanics of RNA translocases. *Methods Enzymol* **511**, 131-147.
- Dohme, F. & Nierhaus, K. H. (1976). Total reconstitution and assembly of 50 S subunits from Escherichia coli Ribosomes in vitro. *J Mol Biol* **107**, 585-599.
- Dombroski, A. J., Brennan, C. A., Spear, P. & Platt, T. (1988). Site-directed alterations in the ATP-binding domain of rho protein affect its activities as a termination factor. *J Biol Chem* **263**, 18802-18809.
- Dombroski, A. J. & Platt, T. (1988). Structure of rho factor: an RNA-binding domain and a separate region with strong similarity to proven ATP-binding domains. *Proc Natl Acad Sci U S A* **85**, 2538-2542.
- Dreyfus, M. (2009). Killer and protective ribosomes. *Prog Mol Biol Transl Sci* **85**, 423-466.
- Dubey, A. K., Baker, C. S., Romeo, T. & Babitzke, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* **11**, 1579-1587.
- Dumont, S., Cheng, W., Serebrov, V., Beran, R. K., Tinoco, I., Pyle, A. M. & Bustamante, C. (2006). RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature* **439**, 105-108.
- Durand, S., Tomasini, A., Braun, F., Condon, C. & Romby, P. (2015). sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol Rev*.
- Dutta, D., Chalissery, J. & Sen, R. (2008). Transcription termination factor rho prefers catalytically active elongation complexes for releasing RNA. *J Biol Chem* **283**, 20243-20251.
- Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* **146**, 533-543.
- Duval, M., Simonetti, A., Caldelari, I. & Marzi, S. (2015). Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics. *Biochimie* **114**, 18-29.
- Eagles, R. M., Balmori-Melián, E., Beck, D. L., Gardner, R. C. & Forster, R. L. (1994). Characterization of NTPase, RNA-binding and RNA-helicase activities of the cytoplasmic inclusion protein of tamarillo mosaic potyvirus. *Eur J Biochem* **224**, 677-684.
- El Yacoubi, B., Bailly, M. & de Crécy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet* **46**, 69-95.
- Elles, L. M. & Uhlenbeck, O. C. (2008). Mutation of the arginine finger in the active site of Escherichia coli DbpA abolishes ATPase and helicase activity and confers a dominant slow growth phenotype. *Nucleic Acids Res* **36**, 41-50.
- Enemark, E. J. & Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**, 270-275.
- Enemark, E. J. & Joshua-Tor, L. (2008). On helicases and other motor proteins. *Curr Opin Struct Biol* **18**, 243-257.
- Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S. & Nudler, E. (2007). An allosteric path to transcription termination. *Mol Cell* **28**, 991-1001.
- Epshtein, V., Dutta, D., Wade, J. & Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**, 245-249.
- Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O. & Salzberg, S. L. (2000). Prediction of transcription terminators in bacterial genomes. *J Mol Biol* **301**, 27-33.
- Ewen, M. E. & Revel, H. R. (1990). RNA-protein complexes responsible for replication and transcription of the double-stranded RNA bacteriophage phi 6. *Virology* **178**, 509-519.
- Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Current opinion in structural biology* **20**, 313-324.
- Fairman-Williams, M. E. & Jankowsky, E. (2012). Unwinding initiation by the viral RNA helicase NPH-II. *J Mol Biol* **415**, 819-832.
- Feklistov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* **68**, 357-376.
- Figuerola-Bossi, N., Valentini, M., Malleret, L., Fiorini, F. & Bossi, L. (2009). Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes Dev* **23**, 2004-2015.
- Filipe, S. R., Pinho, M. G. & Tomasz, A. (2000). Characterization of the murMN operon involved in the synthesis of branched peptidoglycan peptides in Streptococcus pneumoniae. *J Biol Chem* **275**, 27768-27774.
- Fiorini, F., Boudvillain, M. & Le Hir, H. (2013). Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res* **41**, 2404-2415.
- Fiorini, F., Bagchi, D., Le Hir, H. & Croquette, V. (2015). Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat Commun* **6**, 7581.
- Fraser, C. M., Gocayne, J. D., White, O. & other authors (1995). The minimal gene complement of Mycoplasma genitalium. *Science* **270**, 397-403.
- Fuller-Pace, F. V., Nicol, S. M., Reid, A. D. & Lane, D. P. (1993). DbpA: a DEAD box protein specifically activated by 23S rRNA. *EMBO J* **12**, 3619-3626.
- Gai, D., Zhao, R., Li, D., Finkielstein, C. V. & Chen, X. S. (2004). Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**, 47-60.

- Galletto, R. & Tomko, E. J. (2013). Translocation of *Saccharomyces cerevisiae* Pif1 helicase monomers on single-stranded DNA. *Nucleic Acids Res* **41**, 4613-4627.
- Galloway, J. L. & Platt, T. (1988). Signals sufficient for rho-dependent transcription termination at trp t' span a region centered 60 base pairs upstream of the earliest 3' end point. *J Biol Chem* **263**, 1761-1767.
- García-García, C., Frieda, K. L., Feoktistova, K., Fraser, C. S. & Block, S. M. (2015). RNA BIOCHEMISTRY. Factor-dependent processivity in human eIF4A DEAD-box helicase. *Science* **348**, 1486-1488.
- Gegenheimer, P. & Apirion, D. (1981). Processing of procaryotic ribonucleic acid. *Microbiol Rev* **45**, 502-541.
- Geiselmann, J., Yager, T. D., Gill, S. C., Calmettes, P. & von Hippel, P. H. (1992). Physical properties of the *Escherichia coli* transcription termination factor rho. 1. Association states and geometry of the rho hexamer. *Biochemistry* **31**, 111-121.
- Geissmann, T., Marzi, S. & Romby, P. (2009). The role of mRNA structure in translational control in bacteria. *RNA Biol* **6**, 153-160.
- Geissmann, T. A. & Touati, D. (2004). Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J* **23**, 396-405.
- Giuliodori, A. M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., Gualerzi, C. O. & Pon, C. L. (2010). The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol Cell* **37**, 21-33.
- Gocheva, V., Le Gall, A., Boudvillain, M., Margeat, E. & Nollmann, M. (2015). Direct observation of the translocation mechanism of transcription termination factor Rho. *Nucleic Acids Res* **43**, 2367-2377.
- Godoy, T. D., Mian, G. F., Zanollo, R., Yuhara, T. Y., Faria, F. C. & Figueiredo, H. C. P. (2008). Patterns of resistance to florfenicol and bicyclomycin in Brazilian strains of motile aeromonads. *Aquaculture* **285**, 255-259.
- Gogol, E. P., Seifried, S. E. & von Hippel, P. H. (1991). Structure and assembly of the *Escherichia coli* transcription termination factor rho and its interaction with RNA. I. Cryoelectron microscopic studies. *J Mol Biol* **221**, 1127-1138.
- Gong, F. & Yanofsky, C. (2003). Rho's role in transcription attenuation in the tna operon of *E. coli*. *Methods Enzymol* **371**, 383-391.
- Gorbalenya, A. E., Blinov, V. M., Donchenko, A. P. & Koonin, E. V. (1989). An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. *J Mol Evol* **28**, 256-268.
- Gorbalenya, A. E. & Koonin, E. V. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology* **3**, 419-429.
- Gottshall, E. Y., Seebart, C., Gatlin, J. C. & Ward, N. L. (2014). Spatially segregated transcription and translation in cells of the endomembrane-containing bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A* **111**, 11067-11072.
- Gowrishankar, J., Leela, J. K. & Anupama, K. (2013). R-loops in bacterial transcription: their causes and consequences. *Transcription* **4**, 153-157.
- Graves, E. T., Duboc, C., Fan, J., Stransky, F., Leroux-Coyau, M. & Strick, T. R. (2015). A dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and fluorescence. *Nat Struct Mol Biol* **22**, 452-457.
- Greive, S. J., Lins, A. F. & von Hippel, P. H. (2005). Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in *Escherichia coli*. *J Biol Chem* **280**, 36397-36408.
- Griffin, J. E., Gawronski, J. D., Dejesus, M. A., Ioerger, T. R., Akerley, B. J. & Sassetti, C. M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**, e1002251.
- Gross, C. H. & Shuman, S. (1996). Vaccinia virions lacking the RNA helicase nucleoside triphosphate phosphohydrolase II are defective in early transcription. *J Virol* **70**, 8549-8557.
- Gu, M. & Rice, C. M. (2010). Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc Natl Acad Sci U S A* **107**, 521-528.
- Guajardo, R. & Sousa, R. (1997). A model for the mechanism of polymerase translocation. *J Mol Biol* **265**, 8-19.
- Guarente, L. P. & Beckwith, J. (1978). Mutant RNA polymerase of *Escherichia coli* terminates transcription in strains making defective rho factor. *Proc Natl Acad Sci U S A* **75**, 294-297.
- Guell, M., van Noort, V., Yus, E. & other authors (2009). Transcriptome complexity in a genome-reduced bacterium. *Science* **326**, 1268-1271.
- Guerin, M., Robichon, N., Geiselmann, J. & Rahmouni, A. R. (1998). A simple polypyrimidine repeat acts as an artificial Rho-dependent terminator in vivo and in vitro. *Nucleic Acids Res* **26**, 4895-4900.
- Gusarov, I. & Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Mol Cell* **3**, 495-504.
- Gutierrez, P., Kozlov, G., Gabrielli, L., Elias, D., Osborne, M. J., Gallouzi, I. E. & Gehring, K. (2007). Solution structure of YaeO, a Rho-specific inhibitor of transcription termination. *J Biol Chem* **282**, 23348-23353.

- Halls, C., Mohr, S., Del Campo, M., Yang, Q., Jankowsky, E. & Lambowitz, A. M. (2007). Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J Mol Biol* **365**, 835-855.
- Harinarayanan, R. & Gowrishankar, J. (2003). Host factor titration by chromosomal R-loops as a mechanism for runaway plasmid replication in transcription termination-defective mutants of Escherichia coli. *J Mol Biol* **332**, 31-46.
- Hein, P. P., Palangat, M. & Landick, R. (2011). RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. *Biochemistry* **50**, 7002-7014.
- Hein, P. P., Kolb, K. E., Windgassen, T., Bellecourt, M. J., Darst, S. A., Mooney, R. A. & Landick, R. (2014). RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* **21**, 794-802.
- Herbert, K. M., Zhou, J., Mooney, R. A., Porta, A. L., Landick, R. & Block, S. M. (2010). E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J Mol Biol* **399**, 17-30.
- Hershey, J. W., Sonenberg, N. & Mathews, M. B. (2012). Principles of translational control: an overview. *Cold Spring Harb Perspect Biol* **4**.
- Hilbert, M., Karow, A. R. & Klostermeier, D. (2009). The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol Chem* **390**, 1237-1250.
- Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E. & Groisman, E. A. (2012). Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci U S A* **109**, 5376-5381.
- Hollands, K., Sevostyanova, A. & Groisman, E. A. (2014). Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator. *Proc Natl Acad Sci U S A* **111**, E1999-2007.
- Holmqvist, E. & Vogel, J. (2013). A small RNA serving both the Hfq and CsrA regulons. *Genes Dev* **27**, 1073-1078.
- Hopfner, K. P. & Michaelis, J. (2007). Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. *Curr Opin Struct Biol* **17**, 87-95.
- Hori, H. a. T. C. a. H. A. a. T. Y. a. T. K. a. U. T. a. W. K. (2001). Transfer RNA Synthesis and Regulation.
- Howan, K., Smith, A. J., Westblade, L. F., Joly, N., Grange, W., Zorman, S., Darst, S. A., Savery, N. J. & Strick, T. R. (2012). Initiation of transcription-coupled repair characterized at single-molecule resolution. *Nature* **490**, 431-434.
- Howan, K., Monnet, J., Fan, J. & Strick, T. R. (2014). Stopped in its tracks: the RNA polymerase molecular motor as a robust sensor of DNA damage. *DNA Repair (Amst)* **20**, 49-57.
- Hsieh, P. K., Richards, J., Liu, Q. & Belasco, J. G. (2013). Specificity of RppH-dependent RNA degradation in Bacillus subtilis. *Proc Natl Acad Sci U S A* **110**, 8864-8869.
- Hussein, R., Lee, T. Y. & Lim, H. N. (2015). Quantitative characterization of gene regulation by Rho dependent transcription termination. *Biochim Biophys Acta* **1849**, 940-954.
- Ingham, C. J., Hunter, I. S. & Smith, M. C. (1996). Isolation and sequencing of the rho gene from Streptomyces lividans ZX7 and characterization of the RNA-dependent NTPase activity of the overexpressed protein. *J Biol Chem* **271**, 21803-21807.
- Ingham, C. J. (1999). Characterisation of the enzymatic and RNA-binding properties of the Rhodobacter sphaeroides 2.4.1. Rho homologue. *Biochim Biophys Acta* **1446**, 115-125.
- Ingham, C. J., Dennis, J. & Furneaux, P. A. (1999). Autogenous regulation of transcription termination factor Rho and the requirement for Nus factors in Bacillus subtilis. *Mol Microbiol* **31**, 651-663.
- Iost, I., Bizebard, T. & Dreyfus, M. (2013). Functions of DEAD-box proteins in bacteria: current knowledge and pending questions. *Biochim Biophys Acta* **1829**, 866-877.
- Iseki, M., Miyoshi, T., Konomi, T. & Imanaka, H. (1980). Biosynthesis of bicyclomycin. II. Biosynthetic conditions and incorporation of radioactive precursors into bicyclomycin by washed mycelium. *J Antibiot (Tokyo)* **33**, 488-493.
- Italiani, V. C. & Marques, M. V. (2005). The transcription termination factor Rho is essential and autoregulated in Caulobacter crescentus. *J Bacteriol* **187**, 4290-4294.
- Itoh, T. & Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc Natl Acad Sci U S A* **77**, 2450-2454.
- Jagessar, K. L. & Jain, C. (2010). Functional and molecular analysis of Escherichia coli strains lacking multiple DEAD-box helicases. *RNA* **16**, 1386-1392.
- Jain, C. (2008). The E. coli RhlE RNA helicase regulates the function of related RNA helicases during ribosome assembly. *RNA* **14**, 381-389.
- Jankowsky, E. & Fairman, M. E. (2007). RNA helicases--one fold for many functions. *Current opinion in structural biology* **17**, 316-324.
- Jankowsky, E. & Fairman-Williams, M. E. (2010). Chapter 1 An Introduction to RNA Helicases: Superfamilies, Families, and Major Themes. In RNA Helicases, pp. 1-31: The Royal Society of Chemistry.
- Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. *Trends Biochem Sci* **36**, 19-29.
- Jarmoskaite, I. & Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. *Annu Rev Biochem* **83**, 697-725.

- Jiang, X., Diwa, A. & Belasco, J. G. (2000). Regions of RNase E important for 5'-end-dependent RNA cleavage and autoregulated synthesis. *J Bacteriol* **182**, 2468-2475.
- Jin, D. J., Walter, W. A. & Gross, C. A. (1988). Characterization of the termination phenotypes of rifampicin-resistant mutants. *J Mol Biol* **202**, 245-253.
- Jin, D. J., Burgess, R. R., Richardson, J. P. & Gross, C. A. (1992). Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc Natl Acad Sci U S A* **89**, 1453-1457.
- Jonas, K., Edwards, A. N., Simm, R., Romeo, T., Römmling, U. & Melefors, O. (2008). The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Mol Microbiol* **70**, 236-257.
- Jones, P. G., Mitta, M., Kim, Y., Jiang, W. & Inouye, M. (1996). Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in Escherichia coli. *Proc Natl Acad Sci U S A* **93**, 76-80.
- Jousselin, A., Metzinger, L. & Felden, B. (2009). On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol* **17**, 399-405.
- Jäger, S., Fuhrmann, O., Heck, C., Hebermehl, M., Schiltz, E., Rauhut, R. & Klug, G. (2001). An mRNA degrading complex in Rhodobacter capsulatus. *Nucleic Acids Res* **29**, 4581-4588.
- Jäger, S., Hebermehl, M., Schiltz, E. & Klug, G. (2004). Composition and activity of the Rhodobacter capsulatus degradosome vary under different oxygen concentrations. *J Mol Microbiol Biotechnol* **7**, 148-154.
- Jørgensen, M. G., Thomason, M. K., Havelund, J., Valentin-Hansen, P. & Storz, G. (2013). Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev* **27**, 1132-1145.
- Kaberdin, V. R. & Bizebard, T. (2005). Characterization of Aquifex aeolicus RNase E/G. *Biochem Biophys Res Commun* **327**, 382-392.
- Kaberdin, V. R., Singh, D. & Lin-Chao, S. (2011). Composition and conservation of the mRNA-degrading machinery in bacteria. *J Biomed Sci* **18**, 23.
- Kaczanowska, M. & Rydén-Aulin, M. (2007). Ribosome biogenesis and the translation process in Escherichia coli. *Microbiol Mol Biol Rev* **71**, 477-494.
- Kainov, D. E., Pirttimaa, M., Tuma, R., Butcher, S. J., Thomas, G. J., Bamford, D. H. & Makeyev, E. V. (2003). RNA packaging device of double-stranded RNA bacteriophages, possibly as simple as hexamer of P4 protein. *J Biol Chem* **278**, 48084-48091.
- Kainov, D. E., Tuma, R. & Mancini, E. J. (2006). Hexameric molecular motors: P4 packaging ATPase unravels the mechanism. *Cellular and molecular life sciences : CMLS* **63**, 1095-1105.
- Kainov, D. E., Mancini, E. J., Telenius, J., Lissal, J., Grimes, J. M., Bamford, D. H., Stuart, D. I. & Tuma, R. (2008). Structural Basis of Mechanochemical Coupling in a Hexameric Molecular Motor. *J Biol Chem* **283**, 3607-3617.
- Kalarickal, N. C., Ranjan, A., Kalyani, B. S., Wal, M. & Sen, R. (2010). A bacterial transcription terminator with inefficient molecular motor action but with a robust transcription termination function. *J Mol Biol* **395**, 966-982.
- Kalyani, B. S., Muteeb, G., Qayyum, M. Z. & Sen, R. (2011). Interaction with the nascent RNA is a prerequisite for the recruitment of Rho to the transcription elongation complex in vitro. *J Mol Biol* **413**, 548-560.
- Karow, A. R. & Klostermeier, D. (2009). A conformational change in the helicase core is necessary but not sufficient for RNA unwinding by the DEAD box helicase YxiN. *Nucleic Acids Res* **37**, 4464-4471.
- Karow, A. R. & Klostermeier, D. (2010). A structural model for the DEAD box helicase YxiN in solution: localization of the RNA binding domain. *J Mol Biol* **402**, 629-637.
- Kassavetis, G. A. & Chamberlin, M. J. (1981). Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. *J Biol Chem* **256**, 2777-2786.
- Khemici, V., Poljak, L., Luisi, B. F. & Carpousis, A. J. (2008). The RNase E of Escherichia coli is a membrane-binding protein. *Mol Microbiol* **70**, 799-813.
- Kim, D. E. & Patel, S. S. (1999). The mechanism of ATP hydrolysis at the noncatalytic sites of the transcription termination factor Rho. *J Biol Chem* **274**, 32667-32671.
- Kim, D. E., Shigesada, K. & Patel, S. S. (1999). Transcription termination factor Rho contains three noncatalytic nucleotide binding sites. *J Biol Chem* **274**, 11623-11628.
- Kim, D. E. & Patel, S. S. (2001). The kinetic pathway of RNA binding to the Escherichia coli transcription termination factor Rho. *J Biol Chem* **276**, 13902-13910.
- Kim, J. L., Morgenstern, K. A., Griffith, J. P., Dwyer, M. D., Thomson, J. A., Murcko, M. A., Lin, C. & Caron, P. R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89-100.
- Kim, J. N. & Breaker, R. R. (2008). Purine sensing by riboswitches. *Biol Cell* **100**, 1-11.
- Kitahara, K. & Suzuki, T. (2009). The ordered transcription of RNA domains is not essential for ribosome biogenesis in Escherichia coli. *Mol Cell* **34**, 760-766.
- Kitai, K., Kashiwazaki, M., Adachi, Y., Kume, T. & Arakawa, A. (1979). In vitro activity of 39 antimicrobial agents against Treponema hyodysenteriae. *Antimicrob Agents Chemother* **15**, 392-395.

- Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. (2000).** rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* **66**, 1328-1333.
- Klein, S., Lorenzo, C., Hoffmann, S. & other authors (2009).** Adaptation of *Pseudomonas aeruginosa* to various conditions includes tRNA-dependent formation of alanyl-phosphatidylglycerol. *Mol Microbiol* **71**, 551-565.
- Klostermeier, D. & Rudolph, M. G. (2009).** A novel dimerization motif in the C-terminal domain of the *Thermus thermophilus* DEAD box helicase Hera confers substantial flexibility. *Nucleic Acids Res* **37**, 421-430.
- Klostermeier, D. (2013).** Rearranging RNA structures at 75°C? toward the molecular mechanism and physiological function of the *thermus thermophilus* DEAD-box helicase hera. *Biopolymers* **99**, 1137-1146.
- Kogoma, T. (1997).** Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* **61**, 212-238.
- Kohn, H. & Widger, W. (2005).** The molecular basis for the mode of action of bicyclomycin. *Curr Drug Targets Infect Disord* **5**, 273-295.
- Kolb, K. E., Hein, P. P. & Landick, R. (2014).** Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J Biol Chem* **289**, 1151-1163.
- Koslover, D. J., Fazal, F. M., Mooney, R. A., Landick, R. & Block, S. M. (2012).** Binding and translocation of termination factor rho studied at the single-molecule level. *J Mol Biol* **423**, 664-676.
- Kossen, K., Karginov, F. V. & Uhlenbeck, O. C. (2002).** The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA. *J Mol Biol* **324**, 625-636.
- Kotlajich, M. V., Hron, D. R., Boudreau, B. A., Sun, Z., Lyubchenko, Y. L. & Landick, R. (2015).** Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife* **4**.
- Kuang, W. F., Lin, Y. C., Jean, F., Huang, Y. W., Tai, C. L., Chen, D. S., Chen, P. J. & Hwang, L. H. (2004).** Hepatitis C virus NS3 RNA helicase activity is modulated by the two domains of NS3 and NS4A. *Biochem Biophys Res Commun* **317**, 211-217.
- Kuhn, E. (2012).** Toward understanding life under subzero conditions: the significance of exploring psychrophilic "cold-shock" proteins. *Astrobiology* **12**, 1078-1086.
- Küpper, H., Sekiya, T., Rosenberg, M., Egan, J. & Landy, A. (1978).** A rho-dependent termination site in the gene coding for tyrosine tRNA su3 of *Escherichia coli*. *Nature* **272**, 423-428.
- Laalami, S., Zig, L. & Putzer, H. (2014).** Initiation of mRNA decay in bacteria. *Cell Mol Life Sci* **71**, 1799-1828.
- Landick, R. (2006).** The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34**, 1062-1066.
- Landick, R., Wade, J. T. & Grainger, D. C. (2015).** H-NS and RNA polymerase: a love-hate relationship? *Curr Opin Microbiol* **24**, 53-59.
- Larson, M. H., Greenleaf, W. J., Landick, R. & Block, S. M. (2008).** Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132**, 971-982.
- Larson, M. H., Mooney, R. A., Peters, J. M. & other authors (2014).** A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042-1047.
- Laursen, B. S., Sørensen, H. P., Mortensen, K. K. & Sperling-Petersen, H. U. (2005).** Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* **69**, 101-123.
- Leela, J. K., Syeda, A. H., Anupama, K. & Gowrishankar, J. (2013).** Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc Natl Acad Sci U S A* **110**, 258-263.
- Lehnik-Habrink, M., Lewis, R. J., Mäder, U. & Stülke, J. (2012).** RNA degradation in *Bacillus subtilis*: an interplay of essential endo- and exoribonucleases. *Mol Microbiol* **84**, 1005-1017.
- Lehnik-Habrink, M., Rempeters, L., Kovács, Á., Wrede, C., Baierlein, C., Krebber, H., Kuipers, O. P. & Stülke, J. (2013).** DEAD-Box RNA helicases in *Bacillus subtilis* have multiple functions and act independently from each other. *J Bacteriol* **195**, 534-544.
- Levin, M. K., Gurjar, M. & Patel, S. S. (2005).** A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. *Nat Struct Mol Biol* **12**, 429-435.
- Li, Z. & Deutscher, M. P. (1996).** Maturation pathways for *E. coli* tRNA precursors: a random multienzyme process in vivo. *Cell* **86**, 503-512.
- Libri, D. (2015).** Sleeping Beauty and the Beast (of pervasive transcription). *RNA* **21**, 678-679.
- Lim, J., Thomas, T. & Cavicchioli, R. (2000).** Low temperature regulated DEAD-box RNA helicase from the Antarctic archaeon, *Methanococcoides burtonii*. *J Mol Biol* **297**, 553-567.
- Lisal, J., Kainov, D. E., Lam, T. T., Emmett, M. R., Wei, H., Gottlieb, P., Marshall, A. G. & Tuma, R. (2006).** Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage. *Virology* **351**, 73-79.
- Liu, F., Putnam, A. & Jankowsky, E. (2008).** ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A* **105**, 20209-20214.

- Liu, X. & Matsumura, P. (1995).** An alternative sigma factor controls transcription of flagellar class-III operons in *Escherichia coli*: gene sequence, overproduction, purification and characterization. *Gene* **164**, 81-84.
- Lowery, C. & Richardson, J. P. (1977).** Characterization of the nucleoside triphosphate phosphohydrolase (ATPase) activity of RNA synthesis termination factor p. II. Influence of synthetic RNA homopolymers and random copolymers on the reaction. *J Biol Chem* **252**, 1381-1385.
- Lowery-Goldhammer, C. & Richardson, J. P. (1974).** An RNA-dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. *Proc Natl Acad Sci U S A* **71**, 2003-2007.
- Ma, W. & Schulten, K. (2015).** Mechanism of Substrate Translocation by a Ring-Shaped ATPase Motor at Millisecond Resolution. *J Am Chem Soc.*
- Magyar, A., Zhang, X., Kohn, H. & Widger, W. R. (1996).** The antibiotic bicyclomycin affects the secondary RNA binding site of *Escherichia coli* transcription termination factor Rho. *J Biol Chem* **271**, 25369-25374.
- Mallam, A. L., Del Campo, M., Gilman, B., Sidote, D. J. & Lambowitz, A. M. (2012).** Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* **490**, 121-125.
- Mancini, E. J., Kainov, D. E., Grimes, J. M., Tuma, R., Bamford, D. H. & Stuart, D. I. (2004).** Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. *Cell* **118**, 743-755.
- Mandal, M. & Breaker, R. R. (2004).** Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463.
- Marcaida, M. J., DePristo, M. A., Chandran, V., Carpousis, A. J. & Luisi, B. F. (2006).** The RNA degradosome: life in the fast lane of adaptive molecular evolution. *Trends Biochem Sci* **31**, 359-365.
- Markkula, A., Mattila, M., Lindström, M. & Korkeala, H. (2012).** Genes encoding putative DEAD-box RNA helicases in *Listeria monocytogenes* EGD-e are needed for growth and motility at 3°C. *Environ Microbiol* **14**, 2223-2232.
- Martin, R., Straub, A. U., Doebele, C. & Bohnsack, M. T. (2013).** DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol* **10**, 4-18.
- Marzi, S., Fechter, P., Chevalier, C., Romby, P. & Geissmann, T. (2008).** RNA switches regulate initiation of translation in bacteria. *Biol Chem* **389**, 585-598.
- Masse, E. & Drolet, M. (1999a).** R-loop-dependent hypernegative supercoiling in *Escherichia coli* topA mutants preferentially occurs at low temperatures and correlates with growth inhibition. *J Mol Biol* **294**, 321-332.
- Masse, E. & Drolet, M. (1999b).** *Escherichia coli* DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J Biol Chem* **274**, 16659-16664.
- Matsumoto, Y., Shigesada, K., Hirano, M. & Imai, M. (1986).** Autogenous regulation of the gene for transcription termination factor rho in *Escherichia coli*: localization and function of its attenuators. *J Bacteriol* **166**, 945-958.
- McGary, K. & Nudler, E. (2013).** RNA polymerase and the ribosome: the close relationship. *Curr Opin Microbiol* **16**, 112-117.
- Menouni, R., Champ, S., Espinosa, L., Boudvillain, M. & Ansaldi, M. (2013).** Transcription termination controls prophage maintenance in *Escherichia coli* genomes. *Proc Natl Acad Sci U S A* **110**, 14414-14419.
- Mercante, J., Edwards, A. N., Dubey, A. K., Babitzke, P. & Romeo, T. (2009).** Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression. *J Mol Biol* **392**, 511-528.
- Michaux, C., Verneuil, N., Hartke, A. & Giard, J. C. (2014).** Physiological roles of small RNA molecules. *Microbiology* **160**, 1007-1019.
- Miller, O. L., Hamkalo, B. A. & Thomas, C. A. (1970).** Visualization of bacterial genes in action. *Science* **169**, 392-395.
- Minias, A. E., Brzostek, A. M., Korycka-Machala, M., Dziadek, B., Minias, P., Rajagopalan, M., Madiraju, M. & Dziadek, J. (2015).** RNase HI Is Essential for Survival of *Mycobacterium smegmatis*. *PLoS One* **10**, e0126260.
- Mischo, H. E., Gómez-González, B., Grzechnik, P., Rondón, A. G., Wei, W., Steinmetz, L., Aguilera, A. & Proudfoot, N. J. (2011).** Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41**, 21-32.
- Mishra, S. & Sen, R. (2015).** N protein from lambdoid phages transforms NusA into an antiterminator by modulating NusA-RNA polymerase flap domain interactions. *Nucleic Acids Res* **43**, 5744-5758.
- Mitra, A., Angamuthu, K., Jayashree, H. V. & Nagaraja, V. (2009).** Occurrence, divergence and evolution of intrinsic terminators across eubacteria. *Genomics* **94**, 110-116.
- Mitra, A., Misquitta, R. & Nagaraja, V. (2014).** *Mycobacterium tuberculosis* Rho Is an NTPase with Distinct Kinetic Properties and a Novel RNA-Binding Subdomain. *PLoS One* **9**, e107474.
- Miwa, Y., Horiguchi, T. & Shigesada, K. (1995).** Structural and functional dissections of transcription termination factor rho by random mutagenesis. *J Mol Biol* **254**, 815-837.
- Miyamura, S., Ogasawara, N., Otsuka, H., Niwayama, S. & Tanaka, H. (1972).** Antibiotic no. 5879, a new water-soluble antibiotic against gram-negative bacteria. *J Antibiot (Tokyo)* **25**, 610-612.
- Miyoshi, T., Miyairi, N., Aoki, H., Kosaka, M. & Sakai, H. (1972).** Bicyclomycin, a new antibiotic. I. Taxonomy, isolation and characterization. *J Antibiot (Tokyo)* **25**, 569-575.

- Miyoshi, T., Iseki, M., Konomi, T. & Imanaka, H. (1980).** Biosynthesis of bicyclomycin. I. Appearance of aerial mycelia negative strains (am-). *J Antibiot (Tokyo)* **33**, 480-487.
- Mohanty, B. K., Maples, V. F. & Kushner, S. R. (2004).** The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol Microbiol* **54**, 905-920.
- Mohanty, B. K. & Kushner, S. R. (2006).** The majority of *Escherichia coli* mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic Acids Res* **34**, 5695-5704.
- Mohanty, B. K. & Kushner, S. R. (2007).** Ribonuclease P processes polycistronic tRNA transcripts in *Escherichia coli* independent of ribonuclease E. *Nucleic Acids Res* **35**, 7614-7625.
- Mohanty, B. K. & Kushner, S. R. (2008).** Rho-independent transcription terminators inhibit RNase P processing of the secG leuU and metT tRNA polycistronic transcripts in *Escherichia coli*. *Nucleic Acids Res* **36**, 364-375.
- Monnet, J., Grange, W., Strick, T. R. & Joly, N. (2013).** Mfd as a central partner of transcription coupled repair. *Transcription* **4**, 109-113.
- Mooney, R. A., Darst, S. A. & Landick, R. (2005).** Sigma and RNA polymerase: an on-again, off-again relationship? *Mol Cell* **20**, 335-345.
- Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z. & Landick, R. (2009a).** Regulator trafficking on bacterial transcription units in vivo. *Mol Cell* **33**, 97-108.
- Mooney, R. A., Schweimer, K., Rosch, P., Gottesman, M. & Landick, R. (2009b).** Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J Mol Biol* **391**, 341-358.
- Morgan, W. D., Bear, D. G. & von Hippel, P. H. (1983).** Rho-dependent termination of transcription. II. Kinetics of mRNA elongation during transcription from the bacteriophage lambda PR promoter. *J Biol Chem* **258**, 9565-9574.
- Motorin, Y. & Helm, M. (2010).** tRNA stabilization by modified nucleotides. *Biochemistry* **49**, 4934-4944.
- Murakami, K. S., Masuda, S. & Darst, S. A. (2002).** Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280-1284.
- Myong, S., Bruno, M. M., Pyle, A. M. & Ha, T. (2007).** Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* **317**, 513-516.
- Nagarajavel, V., Madhusudan, S., Dole, S., Rahmouni, A. R. & Schnetz, K. (2007).** Repression by binding of H-NS within the transcription unit. *J Biol Chem* **282**, 23622-23630.
- Nakagawa, S., Niimura, Y., Miura, K. & Gojobori, T. (2010).** Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci U S A* **107**, 6382-6387.
- Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E. & Altuvia, S. (2009).** A pH-responsive riboregulator. *Genes Dev* **23**, 2650-2662.
- Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. (2003).** Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**, 437-447.
- Newton, W. A., Beckwith, J. R., Zipser, D. & Brenner, S. (1965).** Nonsense mutants and polarity in the lac operon of *Escherichia coli*. *J Mol Biol* **14**, 290-296.
- Nicolas, P., Mader, U., Dervyn, E. & other authors (2012).** Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* **335**, 1103-1106.
- Nierhaus, K. H. (1991).** The assembly of prokaryotic ribosomes. *Biochimie* **73**, 739-755.
- Nishida, M., Mine, Y., Matsubara, T., Goto, S. & Kuwahara, S. (1972).** Bicyclomycin, a new antibiotic. 3. In vitro and in vivo antimicrobial activity. *J Antibiot (Tokyo)* **25**, 582-593.
- Nowatzke, W. L. & Richardson, J. P. (1996).** Characterization of an unusual Rho factor from the high G + C gram-positive bacterium *Micrococcus luteus*. *J Biol Chem* **271**, 742-747.
- Nowatzke, W. L., Burns, C. M. & Richardson, J. P. (1997a).** Function of the novel subdomain in the RNA binding domain of transcription termination factor Rho from *Micrococcus luteus*. *J Biol Chem* **272**, 2207-2211.
- Nowatzke, W. L., Keller, E., Koch, G. & Richardson, J. P. (1997b).** Transcription termination factor Rho is essential for *Micrococcus luteus*. *J Bacteriol* **179**, 5238-5240.
- Nudler, E. (2012).** RNA polymerase backtracking in gene regulation and genome instability. *Cell* **149**, 1438-1445.
- O'Shea, V. L. & Berger, J. M. (2014).** Loading strategies of ring-shaped nucleic acid translocases and helicases. *Curr Opin Struct Biol* **25**, 16-24.
- Oliva, G., Sahr, T. & Buchrieser, C. (2015).** Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS Microbiol Rev* **39**, 331-349.
- Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W. & Darst, S. A. (2003).** Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**, 335-345.
- Opperman, T. & Richardson, J. P. (1994).** Phylogenetic analysis of sequences from diverse bacteria with homology to the *Escherichia coli* rho gene. *J Bacteriol* **176**, 5033-5043.

- Ow, M. C. & Kushner, S. R. (2002).** Initiation of tRNA maturation by RNase E is essential for cell viability in *E. coli*. *Genes Dev* **16**, 1102-1115.
- Owttrim, G. W. (2013).** RNA helicases: diverse roles in prokaryotic response to abiotic stress. *RNA Biol* **10**, 96-110.
- Pandiani, F., Brillard, J., Bornard, I., Michaud, C., Chamot, S., Nguyen-the, C. & Broussolle, V. (2010).** Differential involvement of the five RNA helicases in adaptation of *Bacillus cereus* ATCC 14579 to low growth temperatures. *Appl Environ Microbiol* **76**, 6692-6697.
- Pani, B., Ranjan, A. & Sen, R. (2009).** Interaction surface of bacteriophage P4 protein Psi required for complex formation with the transcription terminator Rho. *J Mol Biol* **389**, 647-660.
- Pannuri, A., Yakhnin, H., Vakulskas, C. A., Edwards, A. N., Babitzke, P. & Romeo, T. (2012).** Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. *J Bacteriol* **194**, 79-89.
- Paoletti, E. & Moss, B. (1974).** Two nucleic acid-dependent nucleoside triphosphate phosphohydrolases from vaccinia virus. Nucleotide substrate and polynucleotide cofactor specificities. *J Biol Chem* **249**, 3281-3286.
- Park, H. G., Zhang, X., Moon, H. S., Zwiefka, A., Cox, K., Gaskell, S. J., Widger, W. R. & Kohn, H. (1995).** Bicyclomycin and dihydrobicyclomycin inhibition kinetics of *Escherichia coli* rho-dependent transcription termination factor ATPase activity. *Arch Biochem Biophys* **323**, 447-454.
- Park, J. S. & Roberts, J. W. (2006).** Role of DNA bubble rewinding in enzymatic transcription termination. *Proc Natl Acad Sci U S A* **103**, 4870-4875.
- Patel, S. S. & Picha, K. M. (2000).** Structure and function of hexameric helicases. *Annu Rev Biochem* **69**, 651-697.
- Peil, L., Virumäe, K. & Remme, J. (2008).** Ribosome assembly in *Escherichia coli* strains lacking the RNA helicase DeaD/CsdA or DbpA. *FEBS J* **275**, 3772-3782.
- Pelechano, V. & Steinmetz, L. M. (2013).** Gene regulation by antisense transcription. *Nat Rev Genet* **14**, 880-893.
- Peters, J. M., Mooney, R. A., Kuan, P. F., Rowland, J. L., Keles, S. & Landick, R. (2009).** Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci U S A* **106**, 15406-15411.
- Peters, J. M., Vangeloff, A. D. & Landick, R. (2011).** Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol* **412**, 793-813.
- Peters, J. M., Mooney, R. A., Grass, J. A., Jessen, E. D., Tran, F. & Landick, R. (2012).** Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev* **26**, 2621-2633.
- Picard, F., Milhem, H., Loubière, P., Laurent, B., Coccagn-Bousquet, M. & Girbal, L. (2012).** Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. *BMC Genomics* **13**, 528.
- Pinkham, J. L. & Platt, T. (1983).** The nucleotide sequence of the rho gene of *E. coli* K-12. *Nucleic Acids Res* **11**, 3531-3545.
- Porrua, O. & Libri, D. (2013).** A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* **20**, 884-891.
- Porrua, O. & Libri, D. (2015).** Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16**, 190-202.
- Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. (2010).** Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504-508.
- Proux, F., Dreyfus, M. & Iost, I. (2011).** Identification of the sites of action of SrmB, a DEAD-box RNA helicase involved in *Escherichia coli* ribosome assembly. *Mol Microbiol* **82**, 300-311.
- Prud'homme-Généreux, A., Beran, R. K., Iost, I., Ramey, C. S., Mackie, G. A. & Simons, R. W. (2004).** Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. *Mol Microbiol* **54**, 1409-1421.
- Py, B., Higgins, C. F., Krisch, H. M. & Carpousis, A. J. (1996).** A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* **381**, 169-172.
- Pyle, A. M. (2008).** Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys* **37**, 317-336.
- Rabhi, M., Rahmouni, R. & Boudvillain, M. (2010a).** Transcription Termination Factor Rho: A Ring-Shaped RNA Helicase from Bacteria. In *RNA Hélicases*, pp. 243-271. Edited by E. Jankowsky: Royal Society of Chemistry.
- Rabhi, M., Tuma, R. & Boudvillain, M. (2010b).** RNA remodeling by hexameric RNA helicases. *RNA biology* **7**, 655-666.
- Rabhi, M., Espéli, O., Schwartz, A., Cayrol, B., Rahmouni, A. R., Arluison, V. & Boudvillain, M. (2011a).** The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J* **30**, 2805-2816.
- Rabhi, M., Gocheva, V., Jacquinet, F., Lee, A., Margeat, E. & Boudvillain, M. (2011b).** Mutagenesis-based evidence for an asymmetric configuration of the ring-shaped transcription termination factor Rho. *J Mol Biol* **405**, 497-518.
- Racle, J., Picard, F., Girbal, L., Coccagn-Bousquet, M. & Hatzimanikatis, V. (2013).** A genome-scale integration and analysis of *Lactococcus lactis* translation data. *PLoS Comput Biol* **9**, e1003240.

- Rahmouni, A. R. & Wells, R. D. (1992).** Direct evidence for the effect of transcription on local DNA supercoiling in vivo. *J Mol Biol* **223**, 131-144.
- Raimann, E., Schmid, B., Stephan, R. & Tasara, T. (2009).** The alternative sigma factor sigma(L) of *L. monocytogenes* promotes growth under diverse environmental stresses. *Foodborne Pathog Dis* **6**, 583-591.
- Rajkumari, K. & Gowrishankar, J. (2001).** In vivo expression from the RpoS-dependent P1 promoter of the osmotically regulated proU operon in *Escherichia coli* and *Salmonella enterica* serovar Typhimurium: activation by rho and hns mutations and by cold stress. *J Bacteriol* **183**, 6543-6550.
- Raney, K. D., Sharma, S. D., Moustafa, I. M. & Cameron, C. E. (2010).** Hepatitis C virus non-structural protein 3 (HCV NS3): a multifunctional antiviral target. *J Biol Chem* **285**, 22725-22731.
- Redder, P., Hausmann, S., Khemici, V., Yasrebi, H. & Linder, P. (2015).** Bacterial versatility requires DEAD-box RNA helicases. *FEMS Microbiol Rev*.
- Remus, D. & Diffley, J. F. (2009).** Eukaryotic DNA replication control: lock and load, then fire. *Curr Opin Cell Biol* **21**, 771-777.
- Revers, F. & García, J. A. (2015).** Molecular biology of potyviruses. *Adv Virus Res* **92**, 101-199.
- Reynolds, N. M., Lazazzera, B. A. & Ibba, M. (2010).** Cellular mechanisms that control mistranslation. *Nat Rev Microbiol* **8**, 849-856.
- Richardson, J. P., Grimley, C. & Lowery, C. (1975).** Transcription termination factor rho activity is altered in *Escherichia coli* with suA gene mutations. *Proc Natl Acad Sci U S A* **72**, 1725-1728.
- Richardson, J. P. (1982a).** Activation of rho protein ATPase requires simultaneous interaction at two kinds of nucleic acid-binding sites. *J Biol Chem* **257**, 5760-5766.
- Richardson, J. P. (1982b).** Activation of rho protein ATPase requires simultaneous interaction at two kinds of nucleic acid-binding sites. *J Biol Chem* **257**, 5760-5766.
- Richardson, J. P. (1991).** Preventing the synthesis of unused transcripts by Rho factor. *Cell* **64**, 1047-1049.
- Richardson, J. P. (2002).** Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta* **1577**, 251-260.
- Richardson, L. V. & Richardson, J. P. (1992).** Cytosine nucleoside inhibition of the ATPase of *Escherichia coli* termination factor rho: evidence for a base specific interaction between rho and RNA. *Nucleic Acids Res* **20**, 5383-5387.
- Robert, F. & Pelletier, J. (2013).** Perturbations of RNA helicases in cancer. *Wiley Interdiscip Rev RNA* **4**, 333-349.
- Roberts, J. W. (1969).** Termination factor for RNA synthesis. *Nature* **224**, 1168-1174.
- Rocak, S. & Linder, P. (2004).** DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol* **5**, 232-241.
- Rocak, S., Emery, B., Tanner, N. K. & Linder, P. (2005).** Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. *Nucleic Acids Res* **33**, 999-1009.
- Rogers, G. W., Richter, N. J. & Merrick, W. C. (1999).** Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J Biol Chem* **274**, 12236-12244.
- Rogers, G. W., Lima, W. F. & Merrick, W. C. (2001).** Further characterization of the helicase activity of eIF4A. Substrate specificity. *J Biol Chem* **276**, 12598-12608.
- Romeo, T., Vakulska, C. A. & Babitzke, P. (2013).** Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ Microbiol* **15**, 313-324.
- Rosana, A. R., Ventakesh, M., Chamot, D., Patterson-Fortin, L. M., Tarassova, O., Espie, G. S. & Owtrim, G. W. (2012).** Inactivation of a low temperature-induced RNA helicase in *Synechocystis* sp. PCC 6803: physiological and morphological consequences. *Plant Cell Physiol* **53**, 646-658.
- Roy, D. & Lieber, M. R. (2009).** G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol* **29**, 3124-3133.
- Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C. & Sonenberg, N. (1990).** Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol Cell Biol* **10**, 1134-1144.
- Rudolph, M. G. & Klostermeier, D. (2009).** The *Thermus thermophilus* DEAD box helicase Hera contains a modified RNA recognition motif domain loosely connected to the helicase core. *RNA* **15**, 1993-2001.
- Rutherford, S. T., Lemke, J. J., Vrentas, C. E., Gaal, T., Ross, W. & Gourse, R. L. (2007).** Effects of DksA, GreA, and GreB on transcription initiation: insights into the mechanisms of factors that bind in the secondary channel of RNA polymerase. *J Mol Biol* **366**, 1243-1257.
- Régnier, P. & Hajnsdorf, E. (2009).** Poly(A)-assisted RNA decay and modulators of RNA stability. *Prog Mol Biol Transl Sci* **85**, 137-185.
- Régnier, P. & Hajnsdorf, E. (2013).** The interplay of Hfq, poly(A) polymerase I and exoribonucleases at the 3' ends of RNAs resulting from Rho-independent termination: A tentative model. *RNA Biol* **10**, 602-609.
- Saecker, R. M., Record, M. T. & Dehaseth, P. L. (2011).** Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol* **412**, 754-771.

- Santangelo, T. J. & Roberts, J. W. (2004).** Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol Cell* **14**, 117-126.
- Santangelo, T. J. & Artsimovitch, I. (2011).** Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol* **9**, 319-329.
- Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003).** Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77-84.
- Scheffner, M., Knippers, R. & Stahl, H. (1989).** RNA unwinding activity of SV40 large T antigen. *Cell* **57**, 955-963.
- Schmidt, M. C. & Chamberlin, M. J. (1984).** Binding of rho factor to Escherichia coli RNA polymerase mediated by nusA protein. *J Biol Chem* **259**, 15000-15002.
- Schneider, D., Gold, L. & Platt, T. (1993).** Selective enrichment of RNA species for tight binding to Escherichia coli rho factor. *FASEB J* **7**, 201-207.
- Schwartz, A., Margeat, E., Rahmouni, A. R. & Boudvillain, M. (2007a).** Transcription termination factor rho can displace streptavidin from biotinylated RNA. *J Biol Chem* **282**, 31469-31476.
- Schwartz, A., Walmacq, C., Rahmouni, A. R. & Boudvillain, M. (2007b).** Noncanonical interactions in the management of RNA structural blocks by the transcription termination rho helicase. *Biochemistry* **46**, 9366-9379.
- Schwartz, A., Rabhi, M., Jacquinet, F., Margeat, E., Rahmouni, A. R. & Boudvillain, M. (2009).** A stepwise 2'-hydroxyl activation mechanism for the bacterial transcription termination factor Rho helicase. *Nat Struct Mol Biol* **16**, 1309-1316.
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. (2006).** Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* **125**, 287-300.
- Serebrov, V., Beran, R. K. & Pyle, A. M. (2009).** Establishing a mechanistic basis for the large kinetic steps of the NS3 helicase. *J Biol Chem* **284**, 2512-2521.
- Shajani, Z., Sykes, M. T. & Williamson, J. R. (2011).** Assembly of bacterial ribosomes. *Annu Rev Biochem* **80**, 501-526.
- Sharma, C. M., Hoffmann, S., Darfeuille, F. & other authors (2010).** The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250-255.
- Sharpe Elles, L. M., Sykes, M. T., Williamson, J. R. & Uhlenbeck, O. C. (2009).** A dominant negative mutant of the E. coli RNA helicase DbpA blocks assembly of the 50S ribosomal subunit. *Nucleic Acids Res* **37**, 6503-6514.
- Shashni, R., Qayyum, M. Z., Vishalini, V., Dey, D. & Sen, R. (2014).** Redundancy of primary RNA-binding functions of the bacterial transcription terminator Rho. *Nucleic Acids Res.*
- Shepherd, J. & Ibba, M. (2015).** Bacterial transfer RNAs. *FEMS Microbiol Rev* **39**, 280-300.
- Shu, C. J. & Zhulin, I. B. (2002).** ANTAR: an RNA-binding domain in transcription antitermination regulatory proteins. *Trends Biochem Sci* **27**, 3-5.
- Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., Klaholz, B. P. & Yusupov, M. (2009).** A structural view of translation initiation in bacteria. *Cell Mol Life Sci* **66**, 423-436.
- Singh, S. S., Singh, N., Bonocora, R. P., Fitzgerald, D. M., Wade, J. T. & Grainger, D. C. (2014).** Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev* **28**, 214-219.
- Singleton, M. R., Dillingham, M. S. & Wigley, D. B. (2007).** Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry* **76**, 23-50.
- Skordalakes, E. & Berger, J. M. (2003).** Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* **114**, 135-146.
- Skordalakes, E., Brogan, A. P., Park, B. S., Kohn, H. & Berger, J. M. (2005).** Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure* **13**, 99-109.
- Skordalakes, E. & Berger, J. M. (2006).** Structural Insights into RNA-Dependent Ring Closure and ATPase Activation by the Rho Termination Factor. *Cell* **127**, 553-564.
- Soares, E., Schwartz, A., Nollmann, M., Margeat, E. & Boudvillain, M. (2014).** The RNA-mediated, asymmetric ring regulatory mechanism of the transcription termination Rho helicase decrypted by time-resolved Nucleotide Analog Interference Probing (trNAIP). *Nucleic Acids Res* **42**, 9270-9284.
- Somerville, C. R. & Ahmed, A. (1979).** Mutants of Escherichia coli defective in the degradation of guanosine 5'-triphosphate, 3'-diphosphate (pppGpp). *Mol Gen Genet* **169**, 315-323.
- Spencer, P. S. & Barral, J. M. (2012).** Genetic code redundancy and its influence on the encoded polypeptides. *Comput Struct Biotechnol J* **1**, e201204006.
- Squires, C. L., Greenblatt, J., Li, J. & Condon, C. (1993).** Ribosomal RNA antitermination in vitro: requirement for Nus factors and one or more unidentified cellular components. *Proc Natl Acad Sci U S A* **90**, 970-974.
- Starosta, A. L., Lassak, J., Jung, K. & Wilson, D. N. (2014).** The bacterial translation stress response. *FEMS Microbiol Rev* **38**, 1172-1201.
- Steimer, L. & Klostermeier, D. (2012).** RNA helicases in infection and disease. *RNA biology* **9**, 751-771.
- Steinmetz, E. J. & Platt, T. (1994).** Evidence supporting a tethered tracking model for helicase activity of Escherichia coli Rho factor. *Proc Natl Acad Sci U S A* **91**, 1401-1405.

- Stevenson, B. S. & Schmidt, T. M. (1998).** Growth rate-dependent accumulation of RNA from plasmid-borne rRNA operons in *Escherichia coli*. *J Bacteriol* **180**, 1970-1972.
- Stewart, V., Landick, R. & Yanofsky, C. (1986).** Rho-dependent transcription termination in the tryptophanase operon leader region of *Escherichia coli* K-12. *J Bacteriol* **166**, 217-223.
- Stitt, B. L. (1988).** *Escherichia coli* transcription termination protein rho has three hydrolytic sites for ATP. *J Biol Chem* **263**, 11130-11137.
- Stitt, B. L. (2001).** *Escherichia coli* transcription termination factor Rho binds and hydrolyzes ATP using a single class of three sites. *Biochemistry* **40**, 2276-2281.
- Svetlov, V. & Nudler, E. (2011).** Clamping the clamp of RNA polymerase. *EMBO J* **30**, 1190-1191.
- Takemoto, N., Tanaka, Y. & Inui, M. (2015).** Rho and RNase play a central role in FMN riboswitch regulation in *Corynebacterium glutamicum*. *Nucleic Acids Res* **43**, 520-529.
- Tanaka, N. & Schwer, B. (2006).** Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function. *Biochemistry* **45**, 6510-6521.
- Tanaka, N., Aronova, A. & Schwer, B. (2007).** Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev* **21**, 2312-2325.
- Tanaka, Y., Teramoto, H. & Inui, M. (2015).** Regulation of the expression of de novo pyrimidine biosynthesis genes in *Corynebacterium glutamicum*. *J Bacteriol*.
- Tehranchi, A. K., Blankschien, M. D., Zhang, Y., Halliday, J. A., Srivatsan, A., Peng, J., Herman, C. & Wang, J. D. (2010).** The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* **141**, 595-605.
- Thomason, M. K., Fontaine, F., De Lay, N. & Storz, G. (2012).** A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* **84**, 17-35.
- Thomsen, N. D. & Berger, J. M. (2008).** Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases. *Mol Microbiol* **69**, 1071-1090.
- Thomsen, N. D. & Berger, J. M. (2009).** Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**, 523-534.
- Tomar, S. K. & Artsimovitch, I. (2013).** NusG-Spt5 proteins-Universal tools for transcription modification and communication. *Chemical reviews* **113**, 8604-8619.
- Toone, W. M., Rudd, K. E. & Friesen, J. D. (1991).** *deaD*, a new *Escherichia coli* gene encoding a presumed ATP-dependent RNA helicase, can suppress a mutation in *rpsB*, the gene encoding ribosomal protein S2. *J Bacteriol* **173**, 3291-3302.
- Tortorici, M. A., Duquerroy, S., Kwok, J. & other authors (2015).** X-ray structure of the pestivirus NS3 helicase and its conformation in solution. *J Virol* **89**, 4356-4371.
- Touloukhonov, I., Artsimovitch, I. & Landick, R. (2001).** Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **292**, 730-733.
- Touloukhonov, I. & Landick, R. (2003).** The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. *Mol Cell* **12**, 1125-1136.
- Tran, L., van Baarsel, J. A., Washburn, R. S., Gottesman, M. E. & Miller, J. H. (2011).** Single-gene deletion mutants of *Escherichia coli* with altered sensitivity to bicyclomycin, an inhibitor of transcription termination factor Rho. *J Bacteriol* **193**, 2229-2235.
- Traub, P. & Nomura, M. (1968).** Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc Natl Acad Sci U S A* **59**, 777-784.
- Tuckerman, J. R., Gonzalez, G. & Gilles-Gonzalez, M. A. (2011).** Cyclic di-GMP activation of polynucleotide phosphorylase signal-dependent RNA processing. *J Mol Biol* **407**, 633-639.
- Vakulskas, C. A., Potts, A. H., Babitzke, P., Ahmer, B. M. & Romeo, T. (2015).** Regulation of Bacterial Virulence by Csr (Rsm) Systems. *Microbiol Mol Biol Rev* **79**, 193-224.
- Van Assche, E., Van Puyvelde, S., Vanderleyden, J. & Steenackers, H. P. (2015).** RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol* **6**, 141.
- Vieu, E. & Rahmouni, A. R. (2004).** Dual role of boxB RNA motif in the mechanisms of termination/antitermination at the lambda trl terminator revealed in vivo. *J Mol Biol* **339**, 1077-1087.
- Vijayan, V., Jain, I. H. & O'Shea, E. K. (2011).** A high resolution map of a cyanobacterial transcriptome. *Genome Biol* **12**, R47.
- Vogel, J. & Luisi, B. F. (2011).** Hfq and its constellation of RNA. *Nat Rev Microbiol* **9**, 578-589.
- von Moeller, H., Basquin, C. & Conti, E. (2009).** The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat Struct Mol Biol* **16**, 247-254.
- Vvedenskaya, I. O., Vahedian-Movahed, H., Bird, J. G., Knoblauch, J. G., Goldman, S. R., Zhang, Y., Ebright, R. H. & Nickels, B. E. (2014).** Transcription. Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344**, 1285-1289.
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982).** Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945-951.

- Walmacq, C., Rahmouni, A. R. & Boudvillain, M. (2004).** Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. *J Mol Biol* **342**, 403-420.
- Walmacq, C., Rahmouni, A. R. & Boudvillain, M. (2006).** Testing the steric exclusion model for hexameric helicases: substrate features that alter RNA-DNA unwinding by the transcription termination factor Rho. *Biochemistry* **45**, 5885-5895.
- Walstrom, K. M., Dozono, J. M., Robic, S. & von Hippel, P. H. (1997a).** Kinetics of the RNA-DNA helicase activity of *Escherichia coli* transcription termination factor rho. 1. Characterization and analysis of the reaction. *Biochemistry* **36**, 7980-7992.
- Walstrom, K. M., Dozono, J. M. & von Hippel, P. H. (1997b).** Kinetics of the RNA-DNA helicase activity of *Escherichia coli* transcription termination factor rho. 2. Processivity, ATP consumption, and RNA binding. *Biochemistry* **36**, 7993-8004.
- Wang, X., Preston, J. F. & Romeo, T. (2004).** The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J Bacteriol* **186**, 2724-2734.
- Wang, X., Dubey, A. K., Suzuki, K., Baker, C. S., Babitzke, P. & Romeo, T. (2005).** CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Mol Microbiol* **56**, 1648-1663.
- Wang, X., Ji, S. C., Jeon, H. J., Lee, Y. & Lim, H. M. (2015).** Two-level inhibition of galK expression by Spot 42: Degradation of mRNA mK2 and enhanced transcription termination before the galK gene. *Proc Natl Acad Sci U S A* **112**, 7581-7586.
- Wang, Y. & von Hippel, P. H. (1993a).** *Escherichia coli* transcription termination factor rho. II. Binding of oligonucleotide cofactors. *J Biol Chem* **268**, 13947-13955.
- Wang, Y. & von Hippel, P. H. (1993b).** *Escherichia coli* transcription termination factor rho. I. ATPase activation by oligonucleotide cofactors. *J Biol Chem* **268**, 13940-13946.
- Washburn, R. S., Marra, A., Bryant, A. P., Rosenberg, M. & Gentry, D. R. (2001).** rho is not essential for viability or virulence in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **45**, 1099-1103.
- Washburn, R. S. & Gottesman, M. E. (2015).** Regulation of Transcription Elongation and Termination. *Biomolecules* **5**, 1063-1078.
- Washio, T., Sasayama, J. & Tomita, M. (1998).** Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res* **26**, 5456-5463.
- Watt, B. & Brown, F. V. (1983).** The in-vitro activity of bicozamycin against anaerobic bacteria of clinical interest. *J Antimicrob Chemother* **12**, 549-553.
- Wei, R. R. & Richardson, J. P. (2001).** Identification of an RNA-binding Site in the ATP binding domain of *Escherichia coli* Rho by H₂O₂/Fe-EDTA cleavage protection studies. *J Biol Chem* **276**, 28380-28387.
- Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. (2013).** Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431-441.
- Wendrich, T. M., Blaha, G., Wilson, D. N., Marahiel, M. A. & Nierhaus, K. H. (2002).** Dissection of the mechanism for the stringent factor RelA. *Mol Cell* **10**, 779-788.
- Wiegshoff, F., Beckering, C. L., Debarbouille, M. & Marahiel, M. A. (2006).** Sigma L is important for cold shock adaptation of *Bacillus subtilis*. *J Bacteriol* **188**, 3130-3133.
- Wilson, D. N. & Nierhaus, K. H. (2007).** The weird and wonderful world of bacterial ribosome regulation. *Crit Rev Biochem Mol Biol* **42**, 187-219.
- Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. (2004).** Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286.
- Wright, J. J. & Hayward, R. S. (1987).** Transcriptional termination at a fully rho-independent site in *Escherichia coli* is prevented by uninterrupted translation of the nascent RNA. *EMBO J* **6**, 1115-1119.
- Yakhnin, A. V. & Babitzke, P. (2002).** NusA-stimulated RNA polymerase pausing and termination participates in the *Bacillus subtilis* trp operon attenuation mechanism invitro. *Proc Natl Acad Sci U S A* **99**, 11067-11072.
- Yakhnin, A. V., Yakhnin, H. & Babitzke, P. (2008).** Function of the *Bacillus subtilis* transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. *Proc Natl Acad Sci U S A* **105**, 16131-16136.
- Yakhnin, A. V., Baker, C. S., Vakulskas, C. A., Yakhnin, H., Berezin, I., Romeo, T. & Babitzke, P. (2013).** CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Mol Microbiol* **87**, 851-866.
- Yakhnin, A. V. & Babitzke, P. (2014).** NusG/Spt5: are there common functions of this ubiquitous transcription elongation factor? *Curr Opin Microbiol* **18**, 68-71.
- Yakhnin, H., Baker, C. S., Berezin, I., Evangelista, M. A., Rassin, A., Romeo, T. & Babitzke, P. (2011).** CsrA represses translation of sdiA, which encodes the N-acylhomoserine-L-lactone receptor of *Escherichia coli*, by binding exclusively within the coding region of sdiA mRNA. *J Bacteriol* **193**, 6162-6170.
- Yang, Q. & Jankowsky, E. (2006).** The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol* **13**, 981-986.

Yang, X., Molimau, S., Doherty, G. P., Johnston, E. B., Marles-Wright, J., Rothnagel, R., Hankamer, B., Lewis, R. J. & Lewis, P. J. (2009). The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO reports* **10**, 997-1002.

Yanofsky, C., Horn, V., Bonner, M. & Stasiowski, S. (1971). Polarity and enzyme functions in mutants of the first three genes of the tryptophan operon of *Escherichia coli*. *Genetics* **69**, 409-433.

Yokota, H., Chujo, Y. A. & Harada, Y. (2013). Single-molecule imaging of the oligomer formation of the nonhexameric *Escherichia coli* UvrD helicase. *Biophys J* **104**, 924-933.

You, Z., Komamura, Y. & Ishimi, Y. (1999). Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity. *Mol Cell Biol* **19**, 8003-8015.

Young, R. & Bremer, H. (1976). Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate. *Biochem J* **160**, 185-194.

Yu, X., Horiguchi, T., Shigesada, K. & Egelman, E. H. (2000). Three-dimensional reconstruction of transcription termination factor rho: orientation of the N-terminal domain and visualization of an RNA-binding site. *J Mol Biol* **299**, 1279-1287.

Zeller, M. E., Csanadi, A., Miczak, A., Rose, T., Bizebard, T. & Kaberdin, V. R. (2007). Quaternary structure and biochemical properties of mycobacterial RNase E/G. *Biochem J* **403**, 207-215.

Zenkin, N., Naryshkina, T., Kuznedelov, K. & Severinov, K. (2006). The mechanism of DNA replication primer synthesis by RNA polymerase. *Nature* **439**, 617-620.

Zha, D., Xu, L., Zhang, H. & Yan, Y. (2014). The two-component GacS-GacA system activates lipA translation by RsmE but not RsmA in *Pseudomonas protegens* Pf-5. *Appl Environ Microbiol* **80**, 6627-6637.

Zheng, C. & Friedman, D. I. (1994). Reduced Rho-dependent transcription termination permits NusA-independent growth of *Escherichia coli*. *Proc Natl Acad Sci U S A* **91**, 7543-7547.

Zhou, Y. N. & Jin, D. J. (1998). The rpoB mutants destabilizing initiation complexes at stringently controlled promoters behave like "stringent" RNA polymerases in *Escherichia coli*. *Proc Natl Acad Sci U S A* **95**, 2908-2913.

Zhu, A. Q. & von Hippel, P. H. (1998). Rho-dependent termination within the trp t' terminator. I. Effects of rho loading and template sequence. *Biochemistry* **37**, 11202-11214.

Zwiefka, A., Kohn, H. & Widger, W. R. (1993). Transcription termination factor rho: the site of bicyclomycin inhibition in *Escherichia coli*. *Biochemistry* **32**, 3564-3570.

Étude de la conservation de la terminaison de la transcription Rho-dépendante au sein de la biodiversité.

Le facteur bactérien Rho est une ARN hélicase toroïdale utilisant l'énergie issue de l'hydrolyse d'ATP pour transloquer le long de brins ARN et dissocier les obstacles moléculaires se trouvant sur son chemin. La fonction majeure de Rho est de provoquer la terminaison de la transcription. Les nombreuses études du facteur Rho d'*E. coli* (EcRho) ont montré que l'enzyme se fixe aux régions nues des transcrits naissants au niveau de sites Rut (Rho utilization) et utilise son activité de translocase ATPase-dépendante pour rattraper l'ARN polymérase et provoquer la dissociation du complexe d'élongation de la transcription. EcRho est impliquée dans 20 à 50 % des événements de terminaison de la transcription chez *E. coli* et contribue à la régulation globale ainsi qu'à la protection du génome bactérien via divers mécanismes complexes. Le spectre d'activité restreint de la Bicyclomycine (BCM) – un antibiotique inhibant la fonction ATPase de Rho – suggère que Rho, bien qu'essentielle chez la plupart des espèces Gram- comme *E. coli*, est superflue chez la plupart des espèces Gram+. Pour vérifier cette hypothèse et mieux comprendre l'importance de Rho au sein de la diversité bactérienne, nous avons évalué la distribution et la conservation phylogénétique de Rho en utilisant les banques publiques de données génomiques et protéomiques. Nous avons observé que ~92% des espèces analysées (représentant la plupart des phyla bactérien) possèdent Rho (ou un gène rho) et que la présence de ce dernier semble être corrélée avec la complexité du génome, du programme de régulation et de l'écosystème de la bactérie. Cette complexité est illustrée par notre découverte que la protéine régulatrice CsrA contrôle la terminaison Rho-dépendante dans la partie 5' UTR de l'opéron pgaABCD dont l'expression est critique pour la formation de biofilms par *E. coli*. En parallèle de ce travail, nous avons testé la proposition récente que le facteur Rho de *Mycobacterium tuberculosis* (MtbRho) opère par un mécanisme atypique indépendant de son activité ATPase, ce qui le rendrait insensible à la BCM. Nos résultats réfutent cette hypothèse et démontrent sans équivoque que MtbRho utilise un mécanisme similaire à celui de EcRho, pouvant être inhibé par la BCM (mais nécessitant des doses particulièrement élevées). L'ensemble de ces travaux apporte de nouvelles informations illuminant le rôle, le mécanisme et l'importance de Rho au sein de la biodiversité et conforte Rho comme cible prometteuse pour le développement de nouveaux antibiotiques.

Mots clés : hélicase; transcription; régulation génique; biodiversité; *Escherichia coli*; *Mycobacterium tuberculosis*; facteur Rho.

Evaluation of the conservation of Rho-dependent transcription termination across the biodiversity.

The bacterial Rho factor is a ring-shaped, hexameric RNA helicase which uses the energy derived from ATP hydrolysis to translocate along RNA strands, disrupting molecular roadblocks in its way. A major function of Rho is to induce termination of transcription. Extensive studies of the prototypical Rho factor from *E. coli* (EcRho) indicate that the enzyme binds naked regions of nascent RNA transcripts at the level of Rut (Rho utilization) sites and, then, uses its ATPase-dependent translocase activity to catch up with RNA polymerase and to trigger dissociation of the transcription elongation complex. EcRho is implicated in 20 to 50 % of all transcription termination events in *E. coli* and contributes to the global regulation and protection of the bacterial genome through various sophisticated mechanisms. The limited spectrum of activity of the antibiotic Bicyclomycin (BCM) – an inhibitor of Rho's ATPase – suggests that Rho, while critical in many Gram- species such as *E. coli*, is dispensable in most Gram+ species. To verify this assumption and better understand the importance of Rho across the biodiversity, we assessed the phylogenetic distribution and conservation of Rho using public 'omics' databases. We found Rho (or a rho ORF) in ~92% of analyzed species (representing most bacterial phyla), its presence seemingly related to the complexity of the bacterial genome, regulatory program, and ecosystem. An illustration of this complexity is our discovery that regulatory protein CsrA mediates Rho-dependent termination in the 5'UTR of the pgaABCD operon whose expression is critical for biofilm formation by *E. coli*. In parallel, we tested the recent proposal that the Rho factor of *Mycobacterium tuberculosis* (MtbRho) operates by an atypical, ATPase-independent mechanism that would make it immune to BCM. Our results rule out this possibility and unambiguously show that MtbRho uses a mechanism similar to that of EcRho, one that can be inhibited by BCM (albeit at unusually high concentrations). Overall, this work provides key, new information to better understand the role, mechanism, and importance of Rho across the bacterial diversity and endorses Rho as a promising target for the development of new antibiotics.

Key words: helicase; transcription; genic regulation; biodiversity; *Escherichia coli*; *Mycobacterium tuberculosis*; Rho factor.